Exploring a Bayesian hierarchical structure within the Behavioural Perspective Model

By

Andrew Rogers

A Thesis Submitted in Fulfilment of the Requirements for the Degree of Doctor of Philosophy of Cardiff University

Marketing and Strategy Section of Cardiff Business School, Cardiff University

Abstract

This thesis focusses on how the behaviour of consumers can be predicted within the Behavioural Perspective Model's (BPM) theoretical framework. The study focuses on three specific area.

First, a complex functional form is created, utilizing the BPM's Informational and Utilitarian reinforcement in combination with behavioural economic, consumer psychology, marketing and seasonal variables.

Second, the text introduces a hierarchical framework to the model. The data are structured as purchases within household and hence the assumption of independence within household purchase is questioned. The hierarchical framework allows the removal of this assumption. Therefore, hierarchical and non-hierarchical models are constructed and compared to investigate this.

Third, the text discusses the Bayesian paradigm and the differences this brings to model estimation versus the more traditional frequentist methods of calculation. The debate between the Bayesian and frequentist paradigms has been prevalent within mathematical and statistical literature for some time and this text is not meant to directly contribute to this literature. However, the text does explore the potential advantages to the subject matter through the exploration of a Bayesian framework for model estimation. Hence, model estimation through a Bayesian framework is employed employing both vague and informed prior distribution, with the informed priors calibrated from frequentist estimates.

ANNEX 1:

Specimen layout for Declaration/Statements page to be included in a thesis.

DECLARATION

This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is being submitted concurrently in candidature for any degree or other award.

Signed ………………………………………………………… (candidate) Date …2/3/18……………

STATEMENT 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of …PhD…

Signed ………………………………………………………… (candidate) Date …2/3/18……………

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated, and the thesis has not been edited by a third party beyond what is permitted by Cardiff University's Policy on the Use of Third Party Editors by Research Degree Students. Other sources are acknowledged by explicit references. The views expressed are my own.

Signed ………………………………………………………… (candidate) Date …2/3/18……………

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available online in the University's Open Access repository and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed ………………………………………………………… (candidate) Date …2/3/18……………

STATEMENT 4: PREVIOUSLY APPROVED BAR ON ACCESS

I hereby give consent for my thesis, if accepted, to be available online in the University's Open Access repository and for inter-library loans **after expiry of a bar on access previously approved by the Academic Standards & Quality Committee.**

Signed ………………………………………………………… (candidate) Date …2/3/18……………

Acknowledgements

I would like to thank my supervisors Professor Gordon Foxall and Dr Peter Morgan for their time, advice and support throughout the PhD.

Thank you is also due to my colleagues in room T37 at Aberconway building, especially Anna Scedrova for the statistics conversations, Hushneara Begum for the support, especially in the early days and Richard Thomas for “scroll-gate” resolution. Special thanks must go to Anna Galaska for making the whole experience one which to look back on with fond memories. A true friend who in the face of grave adversary has achieved so much.

Finally, but by no means least, a huge thanks to Elsie in the PhD office for making the process run so smooth despite the bombardment of questions and requests.

Contents

Chapter 1: Introduction

1.1 Introduction to the study

This text is focused on the field of consumer behaviour. Specifically, the behaviours as can be predicted within the theoretical framework of the Behavioural Perspective Model (BPM). The BPM demonstrates the framework's underlying strength and its agility to be employed to predict various behaviours, spanning multiple situational settings, categories, cultures and geographies (e.g. Foxall and James, 2003; Foxall and Schrezenmaier, 2003; Foxall *et al.*, 2004; Foxall *et al.*, 2006; Oliveira-Castro *et al.*, 2005; Romero *et al.*, 2006). This text builds on this growing research area in three ways.

1. Functional form

Following a literature review and category analysis, a more complex model functional form is proposed encompassing three areas of consumer behaviour. First, the economic behaviour is considered in the form of price elasticity. Second, consumer psychology variables are considered, here in the form of the BPM's Informational and Utilitarian reinforcement. Third, marketing variables are considered, through the lens of the BPM framework. A supermarket own brand indicator is introduced, combined with the Informational and Utilitarian reinforcement variables of the BPM. This allows the understanding of any differences in behaviour associated with the branded products versus the supermarket own products, in the context of the BPM theoretical framework. Also, a seasonality variable is introduced corresponding to the Christmas holiday week. This is due to results seen from the category analysis, which shows a significant reduction in volume during this period. In order to investigate this difference within the BPM framework, an interaction variable is constructed and used versus the Informational and Utilitarian reinforcement variables. This allows the investigation of any changes in behaviour during this period from a consumer psychology perspective.

2. Hierarchical Structure

The second area where this text contributes to the BPM literature is the introduction of a hierarchical modelling framework to the data. The data are constructed as purchases within household, questioning the assumption of independence of behaviour within household. In order to test this, models are constructed of both a non-hierarchical nature (i.e. assuming every transaction is independent which is the usual assumption of a regression based model) and of a hierarchical nature (where the household identifier is used as the underlying hierarchical structure). This framework, is constructed within the BPM framework, again demonstrating the flexibility if the theoretical framework. This means that the assumption of independence is upheld between household but not within household. The hierarchical model is built within the BPM framework.

3. Bayesian Model estimation

The third contribution is the way the models themselves are estimated. The text discusses the Bayesian paradigm and the differences this brings to model estimation versus the more traditional frequentist methods of calculation. The debate between the Bayesian and frequentist paradigms has been prevalent within mathematical and statistical literature for some time and this text is not meant to directly contribute to this literature. However, the text does explore the potential advantages to the subject matter through the exploration of a Bayesian framework for model estimation. Hence, model estimation through a Bayesian framework is used employing both vague and informed prior distribution, with the informed priors calibrated from frequentist estimates. The text will argue that the advantages of using both Bayesian and frequentist tools provide the researcher with a larger analysis tool kit and agrees with Little (2006) on the view that the 21st century should be about pragmatism while utilizing a broad range of methods, including Bayesian and frequentist, in order to furthering consumer behaviour understanding. It also further demonstrates the flexibility of the BPM to estimate model parameters through a Bayesian process.

1.2 Chapter overview of the study

1.2.1 Chapter 2: Literature Review

The literature will focus on four areas. First, the area of consumer psychology is explored through the lens of both a cognitive and behavioural approach. Second, the text will favour the view of the behavioural understanding of consumer psychology over the cognitive and arguments are presented to support this view.

Third, the text introduces the Behavioural Perspective Model (BPM), which is one of the most advanced program of radical consumer behaviour (Wells, 2014). The text describes how the model benefits the field of consumer behaviour psychology and how the model is a pragmatic approach to understanding consumer behaviour. Finally, the concept of Bayesian inference is introduced. It is argued the growth in this paradigm over recent years should not be ignored, with some psychologists arguing these Bayesian methods can be an advantage to the field of consumer psychology by giving additional tools for analysis (Andrews and Baguley, 2013). This is also echoed by other scholars outside the field of consumer psychology (e.g. Little, 2006). The main point of discussion between the Bayesian and frequentist paradigms is the incorporation of a prior distribution within the modelling process. This prior distribution has a direct influence on the parameter inference. This text will make use of both vague prior and informative prior distributions and considers how this affects the parameter inference.

1.2.2 Chapter 3: Data discussion and category review

The data within the study refers to four categories within the Fast Moving Consumer Goods (FMCG), namely biscuits, fruit juice, yellow fats and beans.
The data discussion presents the analysis of the distribution of each category and where necessary, the data is recoded and cleaned, resulting in a data set more appropriate for analysis.

In order to better understand the data, a category analysis is offered for each of the four categories in turn. The resulting analysis offers insights into the economic, BPM and seasonality variables. It also offers seasonality hypotheses to be explored.

1.2.3 Chapter 4: Initial analysis

Following the category review and literature review, exploration of the data is undertaken through the formal univariate statistical analysis of each of the economic, BPM and marketing variables. This helps to formulate the research questions that underpins the study. This also indicates potential relationships between the variables that helps the model build.

This analysis uses frequentist methods of a continuous and categorical nature, depending on the nature of the underlying data.

1.2.4 Chapter 5: Research question construction

Based on a combination of the literature review, category analysis and initial analysis, a series of research questions are constructed and discussed. The research questions are based on the three areas of contribution outlined at the start of this chapter. In brief, they are outlined below, though a much more thorough description is offered within chapter 5 together with a discussion about how the research questions are formulated for category and model specific sub sections of each research question.

RQ1: Does the average price of the products within the category influences consumer economic behaviour?

RQ2: Are the BPM psychological variables accounting for consumer behaviour for each category. the nature of the supermarket own brand impacting consumer behaviour of the category through differing behaviour at a consumer psychological level, either at a utilitarian and/or informational reinforcement level?

RQ3: Is the nature of the supermarket own brand impacting consumer behaviour of the category through differing behaviour at a consumer psychological level, either at a utilitarian and/or informational reinforcement level?

RQ4: Is the seasonal Christmas week impacting consumer behaviour within the category, through various levels of utilitarian and/or informational reinforcement during the Christmas seasonal week?

RQ5: Will the modelling of the biscuits category within the BPM structure benefit from a hierarchical model structure? What differences in interpretation would be included versus a non-hierarchical framework?

RQ6: How will Bayesian inference utilizing informative and vague priors impact the predictive nature of the model and the interpretation of the parameters?

RQ7: Does a combined category model, incorporating all four categories in one model, utilising a pooled parameter structure help the interpretation of consumer behaviour both from a model diagnostic and interpretation perspective? Or does a combined category model, incorporating all four categories in one model, utilising an offset parameter structure help the interpretation of consumer behaviour both from a model diagnostic and interpretation perspective?

RQ8: How does the diagnostic measured and parameter estimation differ between treating the data as four separate category models versus one combined cross-category model.

1.2.5 Chapter 6: Methods

This chapter builds on the knowledge gained from previous studies and describes the methods that are be used to construct the subsequent statistical models together with their analysis and interpretation.

The models are initially built as four separate category entities. From the literature review, each category comprises of three model builds, comprising of a non-hierarchical model, a hierarchical model with vague priors and a hierarchical model with informative priors.
The methods chapter explains how the variables are constructed and interpreted in terms of the functional form of the model. An important aspect of Bayesian inference is the prior distribution and the way in which the prior distributions are constructed; hence, this is also addressed.

As discussed in the literature review, the concept of a hierarchical model is introduced next. This model structure removes the assumption of independence amongst household. The changes required to the functional form of the non-hierarchical model are explained. The prior variance terms for a hierarchical model are also addressed. This results in three functional forms for each of the category models, namely non-hierarchical, hierarchical with vague priors and hierarchical with informative priors.

Next the prospect of a combined model is discussed whereby all categories are modelled simultaneously. Previously, the four categories were modelled as separate entities has an underlying assumption of independence in terms of how the categories are purchased from an economic and psychological behavioural perspective. By including the categories within one combined model removes this assumption of independence between categories since there is a common household identifier running across categories.

The combined model can be represented as a pooled structure whereby one coefficient estimate is present for each category variable (e.g. one coefficient for price representing all four categories). Alternatively, the combined model can be built whereby each category has its own coefficient for each variable (e.g. four individual coefficients for price, representing each of the four categories). If a fixed effects model is utilised then the further question arises, whether an offset approach is used whereby a category is chosen as the *base category* and the other category coefficients are *offset* to this. Another possibility is that each category has its own specific estimate for the variable in question. The benefits and limitations are discussed for each and an argument presented for the offset methodology.

The interpretation of the coefficients will vary depending on whether a pooled model or fixed effects model is utilised and this interpretation is explained. The complexity increases with multiple category models.

Finally, an overview of the Bayesian modelling process is presented together with how the model diagnostics are to be interpreted both from a Bayesian perspective and a frequentist perspective (both paradigms included as discussed in the literature).

1.2.6 Chapter 7: Separate Category Analysis

The four models are constructed and run in turn using the methodology described in previous chapter. For each category in turn, the model diagnostics are discussed and compared between the three models (non-hierarchical, hierarchical with vague priors and hierarchical with informative priors). In general, it concludes the hierarchical models are a better representation (statistically) of the underlying data though there is little between the prior and informative models. This is due to largely agreement between the prior distributions and the likelihood from the data.

Next the coefficients for each of the variables of each model are discussed. The results will show the choice of whether the model is run with a hierarchical or non-hierarchical structure can have a bearing on how results are interpreted. Also, the nature of the prior (vague vs. informative) also has a bearing. This underlines the importance of choice of both functional form and prior distribution during model build.

1.2.7 Chapter 8: Combined category analysis

This section builds on the methodology discussed in the methods chapter in building a combined model across all four categories. The model uses a non-hierarchical model and a hierarchical model with vague priors (omitting the hierarchical model with informative priors for reasons discussed in the methods section). The models are run as a pooled and fixed effect functional form, estimated using Bayesian inference through MCMC simulation, as per the methods chapter.

The model diagnostics and parameter estimates are discussed and compared for the hierarchical and non-hierarchical models within the pooled structure and then again, within the fixed effects structure.

Next, the model performance and coefficient estimates are compared between the two pooled and two fixed effects models. It is shown that the hierarchical structure is deemed to be the more important factor in terms of model performance; however, an argument is made to support a preference for the fixed effects model over the pooled model, despite little difference statistically (at least in this study of four categories).

Finally, a comparison is offered as to the difference and similarities between the (preferred) hierarchical fixed effects model and the hierarchical model of the four separate category models that were estimated in chapter 7. It is noted there are agreements in direction in terms of the parameter estimates in most cases.

1.2.8 Chapter 9: Discussion

The discussion chapter is further divided into sections. First, the RQs are discussed in turn based in the analysis undertaken.

Second, the concept of the incorporation of Bayesian techniques within management is discussed, given the current dominance of the frequentist paradigm. Potential and current issues are discussed both from a literature perspective and also the experiences gained from this current study.

Finally, limitations and future considerations are discussed.

1.3 Contribution of the study

This study contributes to the consumer behaviour literature through the eclectic Behavioural Perspective Model framework, which has been proven to be useful in understanding consumer behaviour in multi categories and multi geographies. The Behavioural Perspective Model (BPM) (Foxall, 1990/2004, 2010) has been used extensively to understand and predict consumer behaviour (e.g. Foxall, 2016a, b, 2017; Foxall and James, 2003; Foxall and Schrezenmaier, 2003; Foxall *et al.*, 2004; Foxall *et al.*, 2006; Oliveira-Castro *et al.*, 2005; Romero *et al.*, 2006). This study increases the understanding of this framework in the following areas discussed below.

From an empirical perspective, the study builds a complex analytical framework incorporating the BPM variables to understand how they affect consumer choice when it comes to supermarket own brands; specifically, understanding the consumer psychology of how these brands are purchased in relation to the nature of their Informational and Utilitarian reinforcement.

Also from an empirical perspective, the seasonal component of the Christmas week is used to assess differences in consumer purchase psychology in terms of Informational and Utilitarian reinforcement within the Christmas week period. This week is selected given the significant difference in volume purchased in this week compared to all other weeks and this is prevalent across all four categories.

From a theoretical perspective, the study introduces a mixed effects hierarchical structure to the model that better resembles both the consumer purchase pattern and the underlying structure of the data. The results are compared against a non-hierarchical model framework and show that the model with a hierarchical structure better reflects the underlying consumer behaviour theoretically and diagnostically.

A second theoretical advancement is the introduction of a Bayesian inference to estimate the parameters of the BPM. Hence, while building on the demonstrated advantages of a hierarchical framework, two Bayesian hierarchical structures are evaluated and compared, relating them to vague prior and informed prior models, with the informed priors calibrated from frequentist estimates. This shows that the interpretation of the posterior distribution of the parameters can vary when different prior distributions are used and highlights the importance of prior information while utilizing a Bayesian approach. The text will argue that the advantages of using both Bayesian and frequentist tools provide the researcher with a larger analysis tool kit and agree with Little (2006) on the view that the 21st century should be about pragmatism while utilizing a broad range of methods for furthering consumer behaviour.

Chapter 2: Literature Review

This chapter is structured in three sections, discussing relevant literature within each. This is to address the nature of how the study is undertaken and presented.

The study is rooted within the field of consumer psychology; hence, the first literature discussion is around the cognitive and behavioural aspects of the field. Second, literature relating to the importance of the brand is evaluated and the extension to the concept of brand equity. Various viewpoints of brand equity are discussed and this study will argue for a behaviourist viewpoint of brand equity as discussed by Foxall (1999b, 2005) and highlighted through the study of Olivieira-Castro *et al.*, (2008). This behaviourist view has been demonstrated through extensive research studies incorporating the Behavioural Perspective Model (e.g. Foxall and James, 2003; Foxall and Schrezenmaier, 2003; Foxall *et al.*, 2004; Foxall *et al.*, 2006; Oliveira-Castro *et al.*, 2005; Romero *et al.*, 2006). Hence, the Behavioural Perspective Model is used as a theoretical basis and a discussion of the Model basis is presented.

Finally, there has been a significant rise in the use of Bayesian techniques within the field of analytics (Efron, 2005). Bayesian techniques offers a larger analysis tool kit for researchers to utilise and the field of psychology could also benefit from this (Andrews and Baguley, 2013). Also, there has only been one study that has utilised Bayesian inference within the Behavioural Perspective Model (Rogers *et al.*, 2017). Hence, a discussion on Bayesian inference is also conducted. It is noted throughout that even though the viewpoints for and against the Bayesian paradigm are presented, this is not a study that is intended to contribute significantly to that debate, which has predominantly been rooted in the statistical literature. Consequently, the discussion addresses the core points of the discussion and gives reasons for the author's viewpoints and reasons as to why this is utilised in this research.

2.1 Introduction to consumer psychology perspectives

This section outlines two viewpoints of consumer behaviour psychology; a cognitive and a behaviourist viewpoint. A brief critique of each is presented with an emphasis on the behavioural viewpoint.

2.2 Cognitive view

Within the fields of both psychology and economics, consumer behaviour has been dominated by the cognitive paradigm (Foxall, 1987; Foxall, 2003; Foxall *et al.*, 2011; Kassarjian 1982). Its dominance can be considered for many reasons: it is a well-established philosophy with a grounded theoretical framework; behaviour which can be recognised and an analytical framework used to measure the results (Foxall, 1986c).

The cognitive paradigm assumes that prior to a consumer's behaviour there is an antecedent series of mental events which can explain the behaviour (Foxall, 1986c). The consumer "thinks and processes information" (Howard, 1983, p. 96) in much the same manner as an artificial intelligence machine would process information (Estes *et al.,* 1983; Newell and Simon, 1972; Neisser 1967; Skinner, 1985). That is, "we think and then act; we have ideas and then put them into words; we experience feelings and then express them; we intend, decide, and choose to act before acting" (Skinner 1985, p. 291). The process is an entirely logical sequence comprising of information coupling about a product with internal beliefs and attitudes of a consumer, leading to an intention to purchase and then, subsequently, the actual purchase is made (Foxall, 1986c). Therefore, it assumes that a consumer knows what they want, is able to obtain, absorb, process, evaluate and store information which then can be searched and reprocessed for future retrieval (Foxall, 1986c; Foxall, 2003). Stent (1975, p. 1057) says this must imply the existence of "inner man" which transforms the product images into perceptions, processes these perceptions and the resulting product is what Skinner (1985, p. 292) calls a "representation of the world". The individual will then fuse this representation with a cognitively-stored history (Skinner, 1985) resulting is a range of possible options for behaviour. These options create uncertainty resulting in a conflict of interest which the consumer reduces or resolves by making cognitive decisions (Foxall, 2005; Hansen, 1976) which form the antecedents of behaviour such as brand choice (Bettman 1979; McGuire 1976a, b). The environment also contributes to the decision process through a cognitively mediated process, and does not directly influence the purchase decision (Foxall, 1986c). In fact, the direction of action is from the individual onto the environment (Skinner, 1985).

An example of how the cognitive process is used to influence consumer decision process is through advertising. The advertisement triggers a mental reaction which leads to the purchase of the advertised brand (e.g., Colley 1961; Lavidge and Steiner 1961; Atkin 1984; Driver and

Foxall 1984). There have been a series of cognitive based models developed (e.g., Engel, *et al.*, 1995; Howard and Sheth, 1969) with maybe the most well-known utilising attitudes and beliefs to inform consumer choice through the theory of reasoned action (Ajzen and Fishbein, 1980) and the Theory of Planned Behaviour (Ajzen, 1985). These state that the attributes of a product form a psychological bond with a consumer's attitudes and beliefs and, as such, will influence the purchase where the attributes of the product match the attitudes and beliefs of the consumer.

2.3 Behaviourist View

The behaviourist framework of consumer choice differs from that of the cognitive as it assumes the internal process of need, information search, purchase and evaluation is replaced by an external behaviour that can be analysed and predicted (Foxall, 2005). Choice is not assumed to be an internal psychological process, but a consequence of reinforcements within a specific environment (Foxall, 1986a; Foxall 1986b).

Behaviourism was first developed in the early 20th century. It states that behaviour is observable and measurable (Foxall, 1987). American psychologist John B. Watson has been credited as being the father of behavioural techniques and was one of the first people to adopt the discipline within the area of consumer research (Bales, 2009; DiClemente and Hantula, 2003). In 1920 Watson was employed by the Walter J. Thompson organisation. His role was to address how psychology could help understand how advertising could take advantage of the increasing industrial production and national distribution of goods (DiClemente and Hantula, 2003).

Watson's approach relied strictly on that which could be verified within the environment, not cognitively. His philosophy was anti mentalist to the extreme, claiming that mental processes had nothing to do a consumer's behaviour. (Reber *et al.,* 2009). His view is often cited as follows

> *"Give me a dozen healthy infants, well-formed, and my own specified world to bring them up in and I'll guarantee to take any one at random and train him to become any type of specialist I might select – doctor, lawyer, artist, merchant-chief and, yes, even beggar-man and thief, regardless of his talents, penchants, tendencies, abilities, vocations, and race of his ancestors" (Watson, 1930, p. 82).*

While this quote may seem extreme, the inclusion of the, often disregarded, next sentence does put the view much more into the wider perspective and questions the views opposing it. Watson next sentence continues…

> *"I am going beyond my facts and I admit it, but so have the advocates of the contrary and they have been doing it for many thousands of years"* (Watson, 1930, p. 82).

Watson argued that a human is an organism and hence its behaviour and consumption could be controlled through behavioural techniques and emotional responses. Therefore, the role of advertising was not just information distribution but also about controlling consumption (DiClemente and Hantula, 2003). Watson said ''…to get hold of your consumer, or better, to make your consumer react, it is only necessary to confront him with either fundamental or conditioned emotional stimuli.'' (Buckley, 1982, p. 212).

Despite Watson's influence on consumer behaviour, research in this field was not evident until the 1960s when Lindsley (1962) performed laboratory style experiments utilising operant type techniques, whereby the respondent was able to control the brightness of a television via a switch. The brightness was associated with the effectiveness of the advertising, consumption level of advertisements and interest or readership of magazine articles (DiClemente and Hantula, 2003; Wells, 2014).
By 1970, behavioural studies had successfully moved from the laboratory setting to studies focussing on social topics such as waste disposal, energy and disease perception triggering interest from the wider social sciences (Donovan, 2011). They included the reduction of retail theft and inappropriate purchases (e.g. underage cigarette sales), all of which operated within the subject's natural environment and successfully influenced consumers, sellers and marketers alike (DiClemente and Hantula, 2003).

Following the success of the behavioural techniques, a discussion ensued on how the role of behavioural studies should progress. Rothschild and Gaidis (1981) argued the focus of studies should be on the immediate reinforcement and hence immediate change of behaviour, while Peter and Nord (1982) suggested that delayed or intermittent response was also useful to marketers. This resulted in a discussion and further research in the use of classical conditioning techniques in consumer behaviour.

2.3.1 Pavlovian classical Conditioning

> *"[Classical conditioning is] an experimental procedure in which a conditioned stimulus (CS) that is, at the outset, neutral with respect to the unconditioned response (UR) is paired with an unconditioned stimulus (US) that reliably elicits the unconditioned response. After a number of such pairings the CS will elicit, by itself, a conditioned response (CR) very much like the UR". (Wells, 2014, p. 1122)*

The best-known example of classical condition is Pavlov's work. This used the sound of a metronome acting as the conditioned stimulus (CS), food was used as the unconditioned stimulus (US) and salivation as the unconditioned and conditioned responses (UR/CR). The food (US) automatically caused the dogs to salivate causing an unconditioned response (UR). The sound of the metronome was paired with the appearance of the food hence becoming the conditioned response (CS). Eventually the dog responded just to the sound of the metronome by salivating (CR) (Macklin, 1986).

The first marketing academic to trial classic conditioning was Gorn (1982) who used music consumers liked and disliked to condition attitudes to a brand of pen. This experiment became prominent in the development of classical conditioning studies (Wells, 2014). Classical conditioning has also been used within advertisements. For example, Allen and Janiszewski (1989, pp. 39-40) associated music together with the strap line "Now you see it, now you don't" as a (US). The brand is the (CS) which is being consumed by an attractive slim woman which is the (UR) that results in the purchase of the brand, the (CR). Razran (1938) used political statements as the (CS) to predict free meals as the (US) resulting in more agreement to political campaign statements when the meals were shown with them. Allen and Janiszewski (1989) paired the country identities of 'Swedish' and 'Dutch' with positive and negative words and found the nationality paired with the positive words had a more positive attitude towards it. Classic conditioning has also been used in associating models with brands in advertisements (McCracken, 1989; Till *et al.*, 2008). Other examples of the use of classic conditioning is the use of sports presenters paired with sports events and products through the use of advertising (Nord and Peters, 1980), fast food restaurants being associated with sizzling hamburgers, soft drinks brands associated with jingles and cereal associated with sports starts (DiClemente and Hantula, 2003).

However, theoretical and practical issues emerged with this act of classic conditioning. Theoretically, unlike animals, humans came to be aware of the pairing of the conditioned stimulus with the unconditioned response (Wells, 2014) and some studies indicated that when this awareness was apparent, the respondent was more positive (Shimp *et al.*, 1991). Other studies showed the conditioning did not happen until the awareness became apparent (Allen and Janiszewski, 1989). Furthermore, the nature of the tests themselves held the dependent variable as more cognitive than behavioural, i.e. the variable would be the inclination to purchase one brand or another or the attitudinal impact the consumer has on the brand (DiClemente and Hantula, 2003). Peter and Olson (1987, p. 306) suggest that ‘cognitive approaches that attempt to describe the internal mechanisms involved in conditioning processes not only add insight but also help to develop more effective conditioning strategies’. Thus, classic conditioning was moving closer to a cognitive theory and seen to be the “seam” between the two (Anderson, 1986, p. 165). From a practical perspective, studies that were based on behavioural outcomes had mixed results, some resulting in a positive relationship to behaviour (e.g. Gorn (1982), Milliman (1982), McCall and Belmont (1996)) while others recorded no apparent influence on behaviour (e.g. Allen and Madden (1985), Kellaris and Cox (1989)).

2.3.2 Operant Conditioning

Operant conditioning says that the behaviour performed will be related to the consequence of how the behaviour was reinforced or punished previously.

> *“[B]ecause behaviour is conceptualised as operating upon the environment to produce consequences it is known as operant behaviour, the process in which the consequences come to influence the behaviour as operant conditioning, and the behavioural psychology which studies the process as operant psychology”*
> Foxall (2002, pp. 27–28).

Unlike classic conditioning, operant conditioning states that “behaviour is shaped and maintained by its consequences” (Skinner 1972, p. 18). Therefore, instead of training the individual to react to stimuli automatically, the stimulus changes the probability of the individual emitting the operant. If the stimulus is withdrawn, then the probability of emitting the operant decreases until eventually it is extinguished. Extinction may be immediate or lagged (Foxall, 1986c). However, as with classic conditioning, the behaviour is externally controlled with no conscious decisions involved in the behaviour. Mental events such as beliefs, intentions or attitudes play no role in the consumer behaviour (Foxall, 1986c).

The three term contingency operant model that Skinner theorized emphasizes the environment in which behaviour takes place. It also says that behaviour is shaped by events which pre cede and ante cede behaviour (Foxall *et al.*, 2011). The model is shown in Fig 1.

$$S^D \rightarrow R \rightarrow S^R$$

Figure 1: Three term contingency operant model

Here S^D is the stimuli, R the response and S^R the reinforcement within a specific environmental setting. The model invokes a higher (or lower) probability of response to specific stimuli based on the reinforcement received in a similar historic situation. Skinner tested the model in a range of environments including learning, verbal behaviour, clinical interventions, politics, and religion (e.g., Skinner, 1953, 1957).

These consequences are known as *reinforcements* and can be positive, negative or a punishment. A *reinforcer* is a condition where the probability of a future response under similar conditions is increased. This response is then known as an *operant*. Similarly, a *punisher* will decrease the probability of a future response under similar conditions (Foxall, 1986c). It is an aversive reaction to the behaviour (e.g. disappointment in quality of a product) and may result in the extinction of behaviour, e.g. the halt in brand consumption/purchase (Nord and Peter, 1980).

The reinforcement of behaviour can be scheduled, i.e. reinforcement occurs on every certain number of desired behaviours (Wells, 2014). These may be fixed reinforcement whereby the reinforcement is scheduled after a set number of behaviours (e.g. after every 3rd desired behaviour) or may be variable rate schedule whereby the reinforcement occurs on average basis, i.e. with a set overall probability (Nord and Peter, 1980). Consumer choice is influenced by the environment in which the behaviour is performed and the rate of reinforcement or punishment that affects the probability of the behaviour (Foxall, 2003). Therefore, the rate at which a certain behaviour is performed is dependent on the other behavioural options available within that same environment and the pattern of reinforcement and punishment they each would induce (Foxall, 2003). The consumer will consider choice as the rate of which the behaviour is performed given the competing behaviours available to the individual; hence, it is the proportion of times that behaviour A is chosen over, say, behaviour B or C (Hermstein, 1970). Choice is therefore "behaviour in the context of other behaviour" Hermstein (1970, p. 225) with no influence from the mental state of the individual (Foxall, 2003). Empirical studies show evidence of how this theory can explain how competing brands are selected in a given environment (Wells, 2014). Skinner's experiments would highlight this when pigeons pecked at coloured buttons at the same rate as the various colours were distributing food and rats pressed levers at the same rate the levers reinforced behaviour (Hermstein, 1970).

Operant conditioning is not researched to the same level as classical conditioning. Despite theoretical discussions (Nord and Peter 1980; Rothschild and Gaidis 1981; Peter and Nord 1982), there has been little academic research involving the subject within consumer behaviour (Foxall, 1986c). This may be down to the longitudinal nature required for operant conditioning or that classical conditioning is easier associated to the cognitive aspects of psychology, which are more prevalent in the field. Alternatively, it may have been down to the lack of availability of a sophisticated framework at the time (DiClemente and Hantula, 2003).

Radical Behaviourism

"Radical behaviourist paradigm (RBP), is a psychological paradigm whose philosophical stance is the opposite of that inherent in [the cognitive paradigm]" (Foxall, 1986c, p. 398). It claims that behaviour can be explained by variables which are entirely non-cognitive and non-intrapersonal (Foxall, 1986c). Whereas the cognitive psychologist will always attempt to derive a rule based approach to match the observable behaviour to unobservable criteria, this is exactly what the radical behaviourist will avoid (Foxall, 2003). RBP is grounded in operant theory, extrapolated from the experimental work on animals performed by the psychologist B. F. Skinner (Skinner 1938, 1950, 1953, 1957, 1969, 1972, 1974). Skinner (1985) says that the antecedents of the environment together with the histories of the environment and individual govern behaviour and these factors affect the behaviour's rate of response (Foxall, 1986). Whereby the cognitive view suggests that the individual acts on the environment, the behavioural view is that the environment influences the individual through stimuli rather than what the individual observes (Skinner, 1985). These influences are due to past reinforcements and influence the probability of behaviour. These views are built on Watson's view of psychology which is "To predict, given the stimulus, what reaction will take place; or, given the reaction, state what the situation or stimulus is that has caused the reaction" (Watson, 1930, p. 11).

2.3 Cognitive vs Behavioural

The cognitive says that the rat learns from pressing the lever that food appears, implying that if a rat presses a lever that results in the distribution of food, the rat has now *learnt* that this is the case and it is now cognitive knowledge stored in the rat's mind. However, there is no direct evidence that this is the case (Skinner, 1985).

In fact, the cognitive paradigm of consumer behaviour has received much criticism through the evidence of low-correlation performance between pre-behavioural claims and actual observed behaviour (Foxall, 1983, 1984; Porto and Oliveira-Castro, 2013), and for attitudinal data versus actual prediction, with a less reliance on consumer information processing than the cognitive models claim (Wicker, 1969; Foxall 1997; Foxall *et al.,* 2011). Ouellette and Wood (1998) found that for well-practiced purchases, habit alone was a better predictor than cognitive claimed intention. Given the dominance of the positivist epistemology within marketing (Hirschman, 1986; Johnson and Duberley, 2011), where predicting consumer's behaviour is of utmost importance it seems interesting that the cognitive based explanation has prevailed as strongly as is seen (Foxall *et al.,* 2011).

The behavioural view sets the environment at the heart of behaviour (Bagozzi, 2000; Foxall, 2003; Skinner; 1985). A person's behaviour is better predicted by understanding the environment in which the behaviour takes place than the psychological profile of that same consumer (Foxall, 1999a) and changes in the behaviour of a consumer are better explained by the changes in the environment than their psychological profile (Studer, 1973; Wohlwill, 1973). Rather, the consumer behaviour within a specific environment can be predicted even though the actual people may be different. It is the nature of their behaviour that remains consistent to the environment (Barker, 1987; Wicker 1987). Through the environment, the evolution of consumer behaviour advances, much like human behaviour, though on a much slower scale (Wells, 2014). Goldsmith (2004, p. 13) says "[H]umans are animals that have evolved over long periods of time. As such, humans behave much like other animals because they learn and adapt due to their interactions with the environment, and their learned behaviour is analogous to animal behaviour so that it can be modelled (described) mathematically as patterns of responses to environmental stimuli". The role of the environment is also part of the cognitive process though, unlike the behavioural viewpoint, it is mediated cognitively rather than directly influencing the purchase decision (Foxall, 1986c). Skinner (1985) agrees saying when a hand is pulled away from a hot object, the cognitive view implies that the person has observed the environment, processed the information, compared it with information stored mentally and hence remove the hand from the object, i.e. the person has felt the environment and then acted accordingly. However, it is the behaviourist view that the behaviour is similar to that of the evolutionary process. Organisms who fail to pull their hands away from the fire will potentially lose use of them and organisms that have survived over time are those with the use of hands. Similarly, eyes were not created to see; it's the species who evolved with eyes were much more likely to survive than those without (Skinner, 1985). The Behavioural Ecology of Consumption is another operant based behavioural model which sees the consumption behaviour as a form of evolutionary process (Hantula *et al.*, 2001; Rajala and Hantula, 2000; Smith and Hantula, 2003). Search, choice and consumption of products evolve on a longer or shorter term basis depending on their significance to the consumer and seen as functionally the same as foraging (Stephens and Krebs, 1986). Skinner argues another reason to dismiss the cognitive view is the consideration of a human community where verbal communication has evolved. Through his communication, people's behaviour can now be influenced through advice, rules, religion or laws without the individual ever having to cognitively experience the behaviour (Skinner, 1985). A driver knows they need to turn the steering wheel of a car to avoid a collision without learning to experience a collision (Skinner, 1985). Furthermore, models which replicate consumer patterns using environmental variables alone, such as the NBD-Dirichlet have been demonstrated (Ehrenberg 1969, 1972, 1974; Ehrenberg and Goodhardt 1980).

Foxall (1986c, p. 404) criticise the cognitive view further by suggesting inconsistencies between what a respondent may claim versus what she may actually do. He says that "behaviours which belong to different classes (e.g. talking about how one will vote and actually voting) will be consistent only when the contingency of reinforcement applicable to both are functionally equivalent" and the same applies for predicting various consumer behaviour (Davies *et al.*, 2002; Foxall, 1986c).

2.5 Pluralism

Despite the research carried out through the RBP, until as recently as 1987, RBP researchers were either ignored, [the field distorted by the inclusion of cognitive variables such as attitudes, needs or beliefs (e.g., Engel and Blackwell 1982, pp. 240-242), blended environmental impacts with cognitive explanations (e.g., Loudon and Dell Bitta 1983, p. 469)] or miss-defined the reinforcement variables (Foxall, 1986c). This has resulted in a lack of discussion and responses to the paradigms in question. Since "*theory has meaning and significance only within the paradigm wherein it is derived*" (Foxall, 1986c, p. 394), this lack of response can lead only to the restriction of intellectual discussion and advancement on the subject of consumer behaviour (Feyerabend 1975). O'Shaughnessy (1997, p. 682) highlights the 'silliness of assuming there is just one overall explanation of buying behaviour', and Foxall (2001) states that the behavioural aspects of his work have never been 'an attempt to reassert the importance of behavioural psychology to the exclusion of cognitive or other perspectives on consumer choice' (Foxall, 2001, p. 166). Furthermore, it has "*never sought to pursue a behaviourist approach to the exclusion of other perspectives; indeed, the coexistence and interaction of multiple theoretical viewpoints is central to its conception of intellectual development*" (Foxall, 2001, p. 183). The exploration of less well-known behavioural psychological approaches and their application to marketing and consumer behaviour respond well to calls both for a more pluralistic and interdisciplinary culture in consumer research (Marsden and Littler, 1998). Thus far, approaches of blending the psychological and behavioural paradigms have been restricted to a blended approach, mixing both attitudes (cognitive) and classical conditioning (behaviourist). Despite much literature within the medical sector on behavioural-cognitive models or treatments, there is very little within the consumer realm with the exception of problem based consumption such as compulsive buying (Kellet and Bolton, 2009), compulsive hoarding (Frost and Hart, 1996), pathological internet behaviour (Davis, 2001), eating disorders (Decaluwé and Braet, 2005; Fairburn *et al.*, 1999) and drug consumption (Tiffany, 1990).

2.6 Literature Review of the Brand

2.6.1 Brand

The history of branding dates back to ancient Egyptian times when brick makers used symbols to identify their products (Farquar, 1989). In the middle ages a trademark was introduced to products. Bakers of bread used a trademark to guarantee the weight of the bread ensuring sub quality products could be traced back to them and hence trademarks were initially seen as liabilities to the manufacturer (Jones and Morgan, 1994). Over time, however, this became a signal of the product quality and is the association we make with trademarks today. These trademarks also give some legal protection to the manufacturer (Farquhar, 1989).

During the nineteenth century, the industrial revolution brought increased transportation networks plus growth in population in America and Europe. This opened the market to domestic products such as medicines and electrical goods. Increased variety of products gave birth to marketing by giving products appropriate brand names, which needed to be pronounceable, memorable, and descriptive of the product (Hart and Murphy, 1998). One of the first brands to make use of brand associations was that of the Whiskey "Old Smuggler", a name purposely chosen to reflect the quality of the product, since smuggled Whiskey was known to be of a high quality (Farquhar, 1989).

2.6.2 Importance of the Brand

There are a variety of ways in which a brand has been defined, which makes establishing an all-encompassing definition almost impossible, though to better understand the brand it is necessary to attempt to understand the concept from different viewpoints (Wood, 1995). A brand is defined by the American Marketing Association as "the name, term, sign, symbol or design or a combination of them intended to identify the goods and services of one seller or group of sellers and to differentiate them from those of competition", (Kotler et al, 1999, p. 442). However, this definition is sometimes criticised for being too product-oriented as it focuses on the visual aspects of a brand (Wood, 2000). However, Wood (2000) says authors have chosen to use the definition as a basis of their own, e.g. Dibb et al (1997) defines a brand as a name, term, design, symbol *or any other feature* which a seller may use to define their products. Hence this difference allows non-visual properties of a brand to be used to create distinction when they use the words *or any other feature*. (Murphy, 1990, p. 1) defines it as "the product or service of a particular supplier which is differentiated by its name and presentation". Whilst Murphy does not specifically claim these to be tangible or intangible, the word presentation would suggest a more visual impact though the Murphy (1990) does specify the importance of a name that may signal more intangible differentiation. These definitions speak nothing of addressing consumer needs, simply differentiating the brand. Also, there is little to suggest how this may benefit the organisation in terms of increasing brand equity or other consumer measures.

Studies focussing the benefit of a strong brand tend to be grouped into those that speak of a brand from a consumer's perspective and those that speak of brand from a firm's perspective (Wood, 2000). From a consumer's perspective, Ambler (1992), defines a brand as a bundle of attributes, which may be real or illusory, rational or emotional, tangible or invisible. Webster (1994) has a similar take claiming a brand can be seen as a bundle of benefits. These definitions are more consumer based and suggest the attributes of a brand are ones which may bring differentiation, tangible or intangible. Here the importance of the product and the relevance of the product attributes to the consumer are being raised and the concept of the brand is defined as a bundle or attributes (or benefits).

Style and Ambler, (2000) define two approaches of how to define a brand, one of which suggests the brand unifies the targeted elements of the marketing mix in a way that increases the brand values and hence the brand equity. This is a similar view to Ambler and Webster though, instead of the brand being reduced to its attributes, they suggest the brand encompasses the bundle of benefits to create a higher equity than the attributes alone. Therefore, the brand name itself is adding value above and beyond its attributes, which is similar to Farquhar (1989, p. 24) who says a brand "enhances the value of a product beyond its functional purpose". De Chernatony and McDonald (1992) suggest that the brand is the added value over the basic commodity product. However, this does suggest that the brand is only responsible for the non-functional elements of the product which contradicts the definition of Dibb et al (1997), who stated that the brand could differentiate on any feature.

Ambler (1992) other definition of the brand suggests that the brand name is the final piece to be added to the attributes and hence it is an add-on element.

Aaker (1996) says a brand can help enrich the understanding of people's perceptions and attitudes towards the product, guiding communication, effort and helping in creating brand equity. Here, Aaker is shifting the benefit of the brand to the firm as it allows them to benefit by being able to differentiate to increase brand equity. He also suggests this differentiation can create brand equity though the way in which it does this is less clear. Murphy (1990) also recognises the importance of brands to an organisation when he states, "Brands can, over time, become a sort of annuity for their owners as the consumer loyalty and affections they engender ad as a guarantee of future demand and hence of future cash flows.".

Within the literature, the benefit of the brand primarily to the consumer or to the organisation tends to polarise studies and there is little on the benefit to both, (Wood, 2000, p. 666) lists a range of studies shown below that emphasise the benefits the consumer.
Aaker (1991), Bennett (1988), Dibb *et al.,* (1997), Kotler *et al.,* (1996), Watkins (1986) emphasise company benefits while Aaker (1996), Ambler (1992), de Chernatony and McDonald (1992), Goodyear (1993), Keller (1993), Levitt (1962), Murphy (1990), Sheth *et al.,* (1991).

However, Wood (2000) does offer a definition to encompass brand definition from the consumer and firm perspective. She says "A brand is a mechanism for achieving competitive advantage for firms, through differentiation (purpose). The attributes that differentiate a brand provide the customer with satisfaction and benefits for which they are willing to pay (mechanism)" (Wood, 2000, p. 666). She goes on to claim that the firm's competitive advantage is financially based (profit, market share etc.) whereas the consumer's benefit is "real or illusory, rational or emotional, tangible or intangible". However, this does seem to suggest that the consumer benefit is not a benefit to the firm in terms of forming bonds with consumers or increasing brand equity, which Aaker (1996) sees as fundamental in his definition.

There also seems to be a wider organisational benefit of achieving a strong brand; for example, Murphy (1990) says brands with a properly registered trademark can last a lifetime if well looked after and can be a source of great value to manufacturer. It forms a 'pact' between the consumer and manufacturer and therefore it is in the manufacturer's interest to maintain the brand and therefore its relationship with the consumer, (Murphy, 1990). Also, retailers are far more receptive to any line extensions being distributed in their store if it is associated with a strong brand, (de Chernatony and McDonald, 1993). Strong brands make it easier for the owners to borrow capital at a cheaper rate, attracts workers due to the brand reputation, and provides economies of scale for research and development. Strong brands can lead to a manufacturer producing two or more brands that operate in the same market category but appeal to different segments through different positioning, (de Chernatony and McDonald, 1993). Finally, historically there has been a determined focus to build strong brands (e.g. Farquar, 1989; Aaker, 1996; Keller, 1998)

What does seem to be common in the definitions is the need for differentiation around the brand (Aaker, 1996; Kotler et al, 1999; Murphy, 1990; Wood, 2000) and Piercy (1997) says competitive differentiation is about giving the consumer what they want, whilst simultaneously getting the desired results for the organisation and Allen (1989) says, differentiation is key in securing competitive advantage.

The concept of the brand has been with us for many a century and the studies suggest the importance of a strong brand is growing in importance. There appears to be different views, however, on the exact role of the brand and whether it is more pertinent to the consumer, the firm or both. Maybe this will vary by the nature of the category or the sector, however what is consistent is the need for the brand to be strong. It seems a natural progression that a way to measure this brand strength was required. The concept of brand equity came to try and fill this need and the literature continues by looking how this could be achieved.

2.6.3 Brand Equity

Given the importance of a strong brand, there was a move to better understand the value of a brand. In the 1980s, an over-reliance on financial short termism with regard to brands was recognised. To strengthen the longer-term vision of a brand, it was seen that a focus on developmental aspects such as R&D, advertising, training, etc. was required. In 1988, a decision was taken by some companies to include an entry to represent this brand asset in the financial ledger of the organisation (Allen, 1990). With Ranks Hovis McDougall plc being the first in the UK to evaluate all of its brands (Wood, 2000). Much debate ensued over the listing of home-grown and acquired brands as assets on the financial leger since historic accounting where valuations were given to tangible items (Stobart, 1990). Extreme views emerged, representing on one hand, the notion that valuation should be based on past transactions and on the other that it should be based on approximating the current value (Foster, 1989).

The justification was that, unlike other industries, the value of the company did not purely depend on the valuation of tangible assets, but was also dependent these intangible assets. This is highlighted by Simon and Sullivan (1993) who claim that Tobin’s Q (which is the market capitalisation divided by the cost of replacement of tangible items) is usually higher for branded companies. This suggests the brand is not capable of being replaced by the tangible items alone.

Soon after, in 1989 the Marketing Science Institute (MSI) focussed on the importance of measuring and managing the brand. The MSI derived its own definition of brand equity as “the set of associations and behaviour on the part of a brand’s customers, channel members and parent corporation that permits the brand to earn greater volume or greater margins than it could without the brand name” (MSI 1989, cited in Chaudhuri,1995). Both practitioners and academics view brand equity as a measure of the true value of the brand and a source of competitive advantage (Lasser *et al.*, 2005). The importance of brand equity has been emphasised by researchers and advertising executives with some organisations (e.g. Interbrand, Total Research Corporation, Millward Brown) dedicating resources to build tracking systems to monitor and to help managers build brand equity (Baldinger 1990; 1992; Ailawadi et al, 2003).
Barwise (1993) suggests another reason for the creation of the brand equity concept was to convince the financial markets of the value these longer-term measures could bring to the brand over and above pure short-term profits.

Despite agreement on the importance of brand equity, there is less agreement on how it is defined and constructed and the measure of equity can be regarded differently when considered as consumer or financial measures (Wood, 2000). However, future financial systems may be designed whereby the brand is a recognised asset and marketers held responsible for managing it in a specific way and hence understanding both the financial and consumer part of equity will be important to all parties (Wood, 1995).

Farquhar (1989) seems to account for this when he says that equity is the added value endowed by the brand name. The term ‘value’ seems to be intentionally vague as he states that value can be depicted in different ways depending on perspective. To the firm, it is the increased cash flow from being associated with the brand through premium prices or reduced costs. From a trade perspective, higher equity can help gain distribution and from a consumer perspective it can increase the relevance of the brand to the consumer (Farquhar, 1989). He says that strong equity can help improve resilience in difficult times, for example the Budweiser brand was equally strong post the US prohibition years as before prohibition. Another example is the discovery of Benzene in Perrier water in 1990, which resulted in the recall of 160M bottles. A year later, Perrier market share was back to pre-1990 levels (Lane, 2013). However, this general term of equity that Farquhar (1989) uses does not really help to define how equity can be monitored and managed and different perspectives have brought different attempts to do this and the equity construct and management has been a focus for many groups of marketing and brand researchers (Punj and Hillyer, 2004).

2.6.3.1 Behavioural based perspective

Within the behavioural perspective, there are varying definitions of brand equity. Cobb-Walgren *et al.* (1995) claim brand equity is measured based on the consumer's market performance or price premium of a product. The MSI definition, also speaks of equity in terms of sales performance by stating equity “permit[s] the brand to earn greater volume or greater margins than it could without the brand name”.
(Ailawadi et al, 2003, p. 1) also claim equity lead to financial gains and say that brand equity is “the marketing effects or outcomes that accrue to a product with its brand name compared with those that would accrue if the same product did not have the brand name”. They test this through a revenue premium formula which can be realised by a branded versus non-branded product by multiplying the volume and price of the branded and unbranded products and assessing the difference.
These definitions deal with the book value or physical sales performance of brands with higher equity. Barwise (1993) however, says it is extremely difficult to understand the value the brand name is adding and virtually impossible to estimate the value of say ‘Coke’ if the

name didn't exist. The author suggests this extends to a category value and it would be extremely difficult to imagine the value of categories without the existence of prominent brands that invest through the marketing mix in an attempt to increase their own brand equity and sales.

However, other authors also claim higher equity go beyond the immediate short-term sales/financial benefits. Biel (1992) defines brand equity in terms of additional cash flow associated with a brand or service, though also claims that brand equity deals with the value of a brand beyond the physical assets associated with its manufacture or provision. Aaker (1991, p. 15) defines equity as "a set of brand assets and liabilities linked to a brand, name and symbol that add or subtract from the value provided by a product or service to a firm and/or to that firm's customers".
Furthermore, high equity does not necessarily mean higher prices, it is about being credible to the claim and discount brands may achieve high equity if they meet their claims (Erdem and Squait, 1998).

In terms of the BPM, Oliveira-Castro et al. (2006) claims the informational aspect of the model can be regarded as akin to that of the equity associated with the product. Hence the equity in this sense is beyond the functional aspects of the product.

What is common to the equity based literature is that a higher equity is seen as a positive and desirable attribute for a brand which leads to increased revenue and success of a brand.

2.6.4 Double Jeopardy of Marketing

Maybe the largest challenge to the concept of equity comes with the concept of Double Jeopardy (DJ). In 1963, William McPhee observed that comic strips which were read by fewer people, were also liked less by those fewer people. Having identified the same pattern amongst radio presenters he concluded that smaller brands suffered in two ways, fewer buyers and less popularity amongst those fewer buyers. He called this 'double jeopardy'- DJ (Ehrenberg et al, 1990). Extensive research shows a similar pattern being observed more widely across category and geography including media ratings, newspapers, automobiles, oil companies and many consumer packaged goods (Ehrenberg et al, 1990; Colombo and Morrison, 1989; Wright and Sharp, 2001; Ehrenberg and Goodhardt, 2002).

Ehrenberg et al (1990) say that marketing practitioners need to be aware of the effect of DJ on loyalty measures as they need to expect smaller brands' loyalty measures to be smaller than larger brands and not to over react when this is the case. Indeed, Ehrenberg and Goodhardt (2002, p. 2) state that "marketing people not knowing about (DJ) on customer loyalty is like rocket scientists not knowing that the earth is round".

Furthermore, research shows that new product launches attain near-instant loyalty and changes in loyalty are almost wholly accounted for by the DJ effect (Ehrenberg and Goodhardt 2000, Wright and Sharp, 2001). The implication is that brands are not strong or weak in equity, simply large or small in size (Ehrenberg and Goodhardt, 2000; Wright and Sharp, 2001).

Models such as the NBD-Dirichlet model have been extensively used to describe how FMCG goods are purchased, including the DJ phenomenon (Goodhardt et al, 1984; Sharp et al, 2012). The lack of marketing mix variables within the NBD-Dirichlet is due to the assumption the model makes of a stationary marketplace where these variables are already accounted for (Bassi, 2011). This is because the marketing mix in large determines the size of the brand and differences in loyalty are systematic (Ehrenberg et al, 1990). It does not assume the marketing mix variables are absent, simply that within a stationary market, brand volume is unaffected by changes in the marketing mix (Ehrenberg, 1972). However, consumers are still making choices, usually from a repertoire of brands (Ehrenberg, 1972) and certain factors will influence which brand they choose at any one time. The NBD-Dirichlet states that these, on average, will form the predictive nature of the stationary market rather than on individual purchases of the consumer within that market place. Hence, the NBD-Dirichlet describes the pattern of purchase rather than the reason for the individual purchase, i.e. "why one person (or household) generally consumes more toothpaste or soup than others, or somewhat prefers brand j to k or vice versa, is not accounted for by the model and is in fact at this stage still largely unknown" (Goodhardt *et al.*, 1984, p. 638).

There have been instances whereby the Dirichlet model has not predicted shares correctly. Bandyopadhyay *et al.,* (2005) observe lower volume brands from smaller consumer repertoires systematically score better on attitudinal measures than lower volume brands in large consumer repertoires. This may also suggest a further effect above and beyond the DJ effect. Also, Fader and Schmittle (1993) found instances when the NBD-Dirichlet model could not explain market share of excessively high or low volume brands and hence this may suggest there are other factors in play. Furthermore, research by Chaudhuri (1995) shows brand loyalty as a mediating variable in the creation of brand equity which allows both equity and DJ to exist as concepts.

I argue that DJ effect is undisputed in terms of the patterns of purchase as seen with the broad range of studies associated with DJ, and models such as the NBD-Dirichlet can accurately account for a market structure. However, this is a bird's eye view of the category. Models such as the BPM helps to understand the motives of the consumption form a ground level perspective, accounting for how behavioural economics and psychological variables influence the purchase of brands and product within a category. These are the variables which management can influence to shape the brand. The advantage of these models, including the BPM, is the application of a consistent theoretical framework to the interpretation of branding through behavioural economics and through the Utilitarian and Informational psychological reinforcement variables within the BPM (Oliveria-Castro *et al.*, 2008).

2.7 The Behavioural Perspective Model

Behavioural studies have been founded on how rats or pigeons react given set stimuli within a certain environment. Even though this seems a long way from how humans may react to marketing stimuli, the goal is the same, i.e. how do humans react under certain stimuli within a certain environment. Some studies have moved beyond animals to address this through using token economies involving prison inmates, schools and hospitals (Foxall, 2003). The issue is that these do not involve the marketing mix variables, which impact on the everyday consumer, such as price changes, product characteristics, advertising, word of mouth, promotional campaigns, packaging or past experience (Foxall, 2003). Given many of these lie within the marketers control to some extent, they are variables which need to be incorporated in such studies. Also, when more emphasis is given to possible effects of situational variables and to measures of behaviour, the level of prediction of behaviour increases substantially (Foxall, 1997). The development of the Behavioural Perspective Model began a move to a more radical behavioural view (Foxall, 1987). The consumer behaviour analysis (CBA) programme that followed is now the most developed programme of radical behaviourism principles to consumer behaviour (Wells, 2014). It is routed in the intercept of behavioural economics, economic psychology and marketing science and uses behavioural theory to interpret consumer behaviour (Foxall, 2001). Many studies have stemmed from behavioural psychology and consumer behaviour, with one of the earliest being that of the development of the Behavioural Perspective Model (BPM) (Wells, 2014).

The BPM model is designed in such a way to comply with the logical scientific epistemology where observations are intelligible, the resulting knowledge can be replicated in a consistent way, the knowledge can be generalized to a wider population and that it is not subject to intervention of the experimenter (Foxall, 1999a).

The BPM has been used as a theoretical and methodological behavioural framework to explain consumer choice (Foxall and James, 2003; Foxall and Schrezenmaier, 2003; Foxall *et al.*, 2004; Foxall *et al.*, 2006; Oliveira-Castro *et al.*, 2005; Romero *et al.*, 2006). The model (Fig 2) is an extension of the Skinnerian three-term contingency and proposes that behaviour can be viewed as a function of a consumer's learning history within a specific temporal setting together with the benefits (or disbenefits) to be gained from the action (Foxall, 1990/2004).

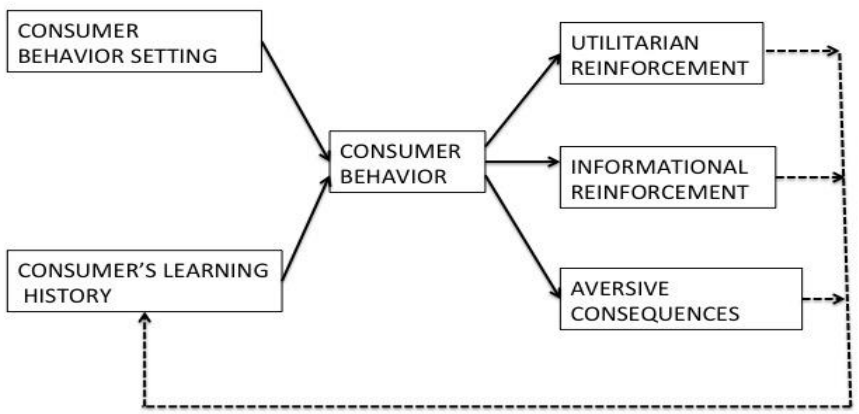

Figure 2: The Behavioural Perspective Model

In line with Skinner's three term contingency, the BPM states that consumption behaviour is followed by a combination of utilitarian and informational reinforcement, and that this *pattern of reinforcement* influences the rate of subsequent behaviour of similar kind (Foxall, 2005). The BPM classifies these reinforcements into two groups. Utilitarian reinforcement is mediated by the product where its attributes and characteristics influence the rate of consumption of the product itself. Utilitarian reinforcements are usually functional, for example consumption of a beverage to quench thirst. Low utilitarian reinforcement usually constitute the basic product whereas increased utilitarian reinforcements usually deliver a functional benefit above this base level, for example within the baked beans category, products which also contain sausage may be seen as a higher utilitarian reinforcement than the plain beans or within the biscuit category, biscuits topped with a chocolate coating may be seen as a higher reinforcement than a plain biscuit (assuming sausages and chocolate are seen as desirable by the consumer of course).

Informational reinforcement is mediated by more social aspects of the brands. Consumers may choose brands with similar utilitarian reinforcement but are deemed to have a higher social value. For example, within FMCG categories, well known brands offer more informational benefits and are seen more as rewards for themselves or family (Foxall *et al.,* 2013). More established brand names can lead to increased informational benefits (Foxall *et al.,* 2013) though social nature of this reinforcement makes it harder to categorise the informational benefits.

Foxall *et al.,* (2004) show that some consumers maximize only the utilitarian reinforcement by purchasing the lower priced products while others maximise their informational reinforcement by purchasing solely premium product, though most consumers purchase a combination of premium and economy products within a category. A further element of the BPM is an aversive consequence, which may result from the behaviour. Often within consumer categories, this may be the monetary compensation required for the consumption; hence the BPM also includes elements of behavioural economics within its framework such as price elasticity coefficients (Foxall, *et al.,* 2011; Oliveira-Castro *et al.,* 2006).

The learning history aspect of the model indicates the experience consumers may have received in a similar behaviour situation and allows the consumer to anticipate the benefits/punishments they may receive under similar situations within the behaviour situation (Foxall, 1999a; Foxall *et al.*, 2011). The learning history alone cannot accurately predict behaviour without placing the consumer's history within a specific behavioural setting. Also, behavioural settings alone offer little predictive power without the consumer's learning history. This is even more the case for relatively open consumer settings (Foxall, 1999a).

The behaviour setting is defined as "the social and physical environment in which the consumer is exposed to stimuli signalling a choice situation" (Foxall *et al.*, 2011, p. 5). Settings range from relatively open (e.g. browsing supermarket shelves where a variety of alternative behaviours are available) to relatively closed (e.g. standing in line in an airport security queue, where a rather inflexible sequence of behaviours is enjoined upon the consumer). Hence, the freedom (in the sense of the number of behavioural options available) the consumer enjoys varies along this open-closed continuum (Foxall, 2013). The consumer behaviour setting includes physical surroundings such as temporal constraints, and social surrounding such as verbal rules (Foxall, 1993). Discriminative stimuli that comprise the consumer behaviour setting include marketing mix variables. Hence, brand and product characteristics are discriminative stimuli that set the occasion for reinforcement, conditional on the consumer's enacting specific purchase and consumption responses (Foxall, 1987). Many consumer situations in relatively affluent communities are relatively open where consumers can freely decide between competitor products (Foxall, 1999a). Choice is not assumed to be an internal psychological process but a consequence of reinforcements within a situational setting (Foxall, 1986a; Foxall 1986b).

2.8 Conclusion

The psychology of consumer behaviour has been dominated by the cognitive paradigm. It has been argued this cognitive process of consumer decision relies on a complex cognitive process. Whilst this has been shown to be a good predictor of planned behaviour, it has limited predictive power to actual behaviour. Radical behaviourist theory, developed on the back of operant based theory, has been proven to be a good representation of actual consumer behaviour. It explains behaviour within a specific environment through rewarding and adversative reinforcements of stimuli. The BPM represents radical behaviourism and has been demonstrated to explain consumer through a number of studies. This study continues by adopting the BPM as a theoretical framework.

2.9 Bayesian inference

2.9.1 Definition

A statistical paradigm, which has become to be known as Bayesian statistics was first published post-humus in 1763 in a work by the Reverend John Bayes entitled “An essay towards solving a problem in the doctrine of chances”. The essay introduced the notion that the probability of an event could be an update of the current view, given the observation of new data. The theorem and its simple proof is illustrated below and “is to the theory of probability what Pythagoras’ theorem is to Geometry” (Jeffreys, 1931, p. 7). Bayes theorem is given mathematically in equation 1.

2.9.2 Bayes Theorem

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}$$

Equation 1: Bayes theorem

2.9.2.1 Proof of Bayes theorem

The probability of A and B happening can be defined as

$$P(A \cap B) = P(A)P(B|A)$$

Similarly, the probability of A and B happening can also be defined as

$$P(A \cap B) = P(B)P(A|B)$$

Hence combining both equations gives the following

$$P(B)P(A|B) = P(A)P(B|A)$$

$$\therefore P(A|B) = \frac{P(A)P(B|A)}{P(B)}$$

Two fundamental issues emerged with the calculation of Bayesian statistics and these forms the basis of debate even today. First, the computational method of calculating the probabilities is difficult and it was 1790 before Pierre-Simon de Laplace demonstrated means by which these could be more easily calculated. Second, the Bayesian paradigm demands prior knowledge of a probability of an event, which is subsequently updated with new information. This requires the existence of this prior knowledge of an event and was seen as biasing the experiment since different prior views could result in differing results (Malakoff, 1999). These two issues are still apparent in discussions today and will be discussed in more detail later in this chapter. Despite these issues, the Bayesian paradigm dominated statistics during the 19th century (Efron, 2005).

Circa 1930, a new methodology was born when Ronald A. Fisher, Egon Pearson and Jerzy Neyman derived a form of probability based on the derivation of inferences for unknown parameters from repeated sampling of a probability distribution (Little, 2006). Under this methodology, the probability of an event is defined as its long run frequency (Koop *et al.*, 2007) and therefore became known as the "frequentist" method. This frequentist method was devoid of a prior knowledge and the mathematics around the calculation of the probability was relatively simple, compared to the Bayesian method (Malakoff, 1999). For these reasons, the frequentist method came to dominate the field of statistics for the 20th century (Efron, 2005; Poirier, 2006). However, the last 30 years has seen a significant increase in Bayesian methods (e.g. Efron, 2005) and this text proceeds to discuss this paradigm and the potential advantages it offers.

2.9.3 Bayesian Inference

The Bayes theorem states the conditional probability of a parameter (θ) given the observed data (X_i) is proportional to the probability of the data given the parameter, multiplied by the probability of the parameter (Congdon, 2003). Or mathematically,

$$P(\theta \mid X_i) \propto P(X_i \mid \theta)P(\theta)$$

This differs from the Bayes theorem itself by omitting the denominator of the right hand side since this is just a normalizing constant (Jeffreys, 1931). The left hand part of the equation is known as the *posterior*- $P(\theta \mid X_i)$ - probability. The terms on the right hand are known as the *likelihood* - $P(X_i \mid \theta)$ - and *prior* - $P(\theta)$ - respectively. The *prior* is the initial belief of a parameter or event before any (new) data is considered. It can come from past studies, expert opinion or what may be considered as common sense (Hansen *et al.,* 2004; O'Hagan, 1994). The *likelihood* is the addition of new data to be evaluated. The *posterior* probability is the blend of both, resulting in an updated view of knowledge based on a combination of the current belief (*prior*) updated by the additional data (*likelihood*).

2.9.4 Differences/Criticisms of the Bayesian Inference approach

2.9.4.1 The Prior Distribution

One fundamental difference between the frequentist and Bayesian paradigms is the explicit inclusion of prior knowledge within the calculation of the posterior distribution. The frequentist would claim this prior information acts as a bias to the experiment since researchers can influence the results by imposing a strong prior distribution on the model. In fact, different results may be obtained from the same data if different researchers choose to apply differing prior distributions (Little, 2006). Efron (2005, p. 1) exemplifies this when observing physicists stating "there's only one way of doing physics, but there seems to be at least two ways to do statistics, and they don't always give the same answers". Another example is risk assessment work by Viscusi (1985) which demonstrates a person's prior knowledge can be systematically biased and, although not criticizing Bayesian philosophy per se, points out the challenges by citing work by Lichtenstein (1978) showing the over assessing of small risks and under assessing of larger risks.

The Bayesian practitioner, however, views this prior knowledge as an important element to the calculation since it matches how a person learns in everyday life (Bernado, 1999). A human mind operates by observing new data and compares this to what (s)he already knows (O' Hagan, 1998). How these pieces "fit together in the light of changing evidence" is fundamentally how the human mind learns (Bernado and Smith, 2000, p. 4). The Bayesian acknowledges the frequentist concern of differing prior distributions leading to differing model estimates though claims this is an issue for the quality of the researchers' knowledge rather than the methods employed to inform the inference (Dunson, 2001). The Bayesian claims frequentist methods themselves are subject to the prior view(s) of researchers being imposed on the model, through the construction of biased questionnaires or leading questions. Leamer (1992) also argues that, in practice, the frequentist researcher must have some prior incline as to the nature of parameters and would reject any absurd model outputs, hence the Bayesian principle is being used in hindsight. Rossi and Allenby (2003) say the fact Bayesian methods require a prior specification is an advantage, since assumptions are explicit and model assumptions in themselves are a form of prior information usually implicit under frequentist based models. Gelman (2010) agrees, quoting Don Rubin when he says scientists interpret uncertainty in a Bayesian manner without realising it, despite working with frequentist methods. (Aspinall, 2010) claims uncertainty should be embraced and quantified, not ignored from the decision making process. O'Hagan (1998, p. 21) agrees saying it is better to embrace and quantify additional information around an experiment and the construction of realistic prior information is better than "relying on ignorance". Researchers are not passive observers and experiments are designed to fit analytic models whether be it within a frequentist or Bayesian framework and the inclusion of the prior is an extension of this build (Efron, 2005).

Dunson (2001) argues the prior distribution can be obtained in a practical manner, deduced from previous studies (hence need not be over complicated) or may be as simple as controlling for absurd results. Practical considerations for both sides of the argument are demonstrated by Efron (2005) in the following example. A drug company performing research may wish to incorporate information from prior studies that can lead to early adoption/rejection of drug development, which they would claim is a better risk for the public and the test subjects of the new drug. However, the FDA would suggest this prior knowledge is of no interest and demand the industry frequentist standards. (Though Efron (2005) notes these standards will have been developed under the dominant frequentist paradigm at the time.)

Gelman, (2010) says the Bayesian paradigm is often discarded as too radical from that of the frequentist, however argues it is the Bayesian that is the more conservative paradigm as it implies the current thinking is preserved unless the data is strong enough to lead to reconsideration. In fact, Gelman (2010, p. 163) strongly criticises frequentist methods claiming "*unbiased estimates and other unregularized classical procedures are noisy and get jerked around by whatever data happen to come by*".

Prior distributions which contain "minimal information" have been used for some time within Bayesian models (Lunn et al, 2012). These are described by Gelman as

"Prior distributions that are uniform, or nearly so, and basically allow the information from the likelihood to be interpreted probabilistically. These are non-informative priors, or maybe, in some cases, weakly informative" (Gelman, 2007). However, Lunn et al (2012) disregard the term non-informative as every prior distribution contains some information and the terms, vague, objective or reference are more suited. The use of these vague priors yield parameter estimates similar to those from maximum likelihood techniques, particularly as the sample size increases and the observed data will have more of a bearing than the prior (Dunson, 2001). Samaniego and Reneau (1994) prefer non-informative prior distribution be used as they mimic a more frequentist approach. Also, Hansen et al (2004) utilize vague priors in their studies.

From a frequentist perspective, it may be argued whether the increased complexity in model computation is necessary for models yielding results similar to frequentist methods. Though from a Bayesian perspective, for such experiments that have no anticipated result, the vague prior is a tool that can reflect this absence of knowledge. This vague prior can be updated for future models of the same form, in light of new information gained from the outputs of the vague prior model and hence laying a baseline for future work (Lunn *et al.*, 2012).

2.9.4.2 Interpretation of the Posterior Distribution

Another area where the two paradigms differ is how the estimated parameter is interpreted. The frequentist views a parameter of a model as unknown but fixed (Abelard, 2012). This means the parameter has a definite value and the analysis is the probability of observing the data given the estimated parameter value (Abelard, 2012) i.e. $P(X_i \mid \theta)$. Recall the Bayesian theorem which states

$$P(\theta \mid X_i) \propto P(X_i \mid \theta)P(\theta)$$

Therefore, Bayes theorem calculates the probability of the parameter, given the data, i.e. $P(\theta \mid X_i)$ (Abelard, 2012). Dunson (2001) claims this is the primary advantage of the Bayesian methodology since the posterior probability is much more intuitive to the layperson than a frequentist p-value since the p-value is the probability of observing a value under a repeated sampling of the null hypothesis. Much more intuitive is the Bayesian interpretation of the posterior estimate as the direct probability of the event occurring (Dunson, 2001). Little (2006, p. 218) agrees saying people would rather have "fixed probability intervals for unknown quantities" (the Bayesian posterior) than "random intervals for fixed quantities" (the frequentist outputs). O'Hagan (1994) says the Bayesian interpretation is more intuitive to management and allows more transparent means of embracing uncertainty of a parameter.

2.9.4.3 Complexity of Calculation

Another issue facing the Bayesian paradigm is the complexity of the calculation. This barrier was identified in the 18th century before Laplace introduced methods for calculating early probability models. The issue is further exacerbated in modern day predictive model building since multi-parameter models require high dimension integration of the posterior function. This is extremely complicated for non-trivial functional form models and initially restricted Bayesian estimation to simple problems (Lunn *et al.,* 2000; Little, 2006). However, the development of the Markov Chain Monte Carlo (MCMC) algorithm in 1995 has all but overcome this issue and has led to strong growth within the Bayesian discipline (Poirier, 2006). This MCMC method allows the posterior to be constructed by the generation of a Monte Carlo style method whereby the shape of the distribution is simulated by a large number of random draws, bypassing the need for the integration of the function (Lunn *et al.,* 2000). This, paired with increased computational power during the same period (which directly facilitates this MCMC methodology) has led to Bayesian models being applied to a broad range of disciplines including astrophysics, genomics, new drug testing, lawsuits, fishing quotas and public policy decisions (Lunn *et al.,* 2000; Malakoff, 1999). So much so, that the Bayesian framework is now seen as a "well established alternative to classical inference" (O'Hagan, 1994, p. 1).

2.9.5 Pragmatism

Little (2006) claims there are three groups of statisticians: frequentists, Bayesians, and pragmatists where pragmatists pick and choose from both frequentist and Bayesian paradigms to suit their analysis needs. Efron (2005) claims models are imperfect in themselves and due diligence is required in checking them; hence, they should not be constrained by a paradigm. The choice of model, functional form, and assumptions around a statistical model will always be incomplete and always contain a degree of uncertainty. As Macdonald (2002, p. 187 [*added by the author for clarity*]) wrote, "if the incompleteness of probability models… were more widely appreciated, psychologists and others might adopt a more reasonable attitude to statistical tests, the debate about statistical inference [*Bayesian and frequentist*] might die down, and the emphasis could shift toward better understanding and presenting data." Little (2006, p. 1) labels the "19th Century as generally Bayesian, the 20th Century as generally frequentist" and suggests "statistics in the 21st Century will require a combination of Bayesian and frequentist ideas". The use of two paradigms increases the number of tools available for researchers to utilize. The concept of pragmatism suggests that both paradigms are used to help the analysis process. Little (2006, 2011) claims that the Bayesian paradigm better lends itself to assert model inference; however, there is a lack of Bayesian tools to assess the model diagnostically. Hence, a natural compromise for the model development and assessment is to incorporate frequentist tests. Andrews and Baguley (2013) claim that the field of psychology needs to use the range of tools provided by both the frequentist and Bayesian methods to help solve the complex problems faced in the field.

A further blend of combining both paradigms is evident from a calibrated Bayes method of model construction (Efron, 2005; Little, 2006). The approach involves deriving estimates of the prior distribution of a Bayesian model by using frequentist methods. Rubin (1984) in Little (2006, p. 7) states, "The applied statistician should be Bayesian in principle and calibrated to the real world in practice." Bayesian models benefit from a thorough model specification encompassing the likelihood and prior dimensions (Little, 2006); however, the models would benefit from the rigorous procedures of model evaluation seen within the frequentist environment (Rubin, 1984). Hansen *et al.*, (2004) make good use of relevant frequentist diagnostics while evaluating the relevance of the Bayesian model and parameter estimates.

The Bayesian methodology has been favoured in this short text, though the wider philosophical view of this study is very much in line with Efron (2005) and Little's (2006) view of a pragmatic approach to building solutions to statistical problems. Gelman (2010, p. 162) also wisely notes the following:

> *"…the key to a good statistical method is not its underlying philosophy or mathematical reasoning, but rather what information the method allows us to use. Good methods make use of more information."*

Two approaches are suggested, utilizing both vague and informative priors. The nature of these approaches will be discussed later. The BPM could benefit from exploiting the flexibility of a Bayesian approach, both in terms of the prior distributions and how the individual estimates are interpreted.

2.10 Summary

The concept of brand has stood the test of time and it seems the importance is increasing. There are multiple definitions of a brand though the consistent views suggest strong brands have benefits for consumers and firms alike. The importance of the brand was highlighted in 1988 when the valuation of brands was included on financial ledgers of organisation. This has led to the MSI declaring the importance of understanding brand equity. Much research followed to better understand and define brand equity. This has resulted in diverging views of how brand equity should be defined, measured and even what the components may be, with the role that brand loyalty plays taking a central focus. The behavioural psychology literature has shown how the consumer choice of brand and product can be predicted given the understanding of economic variables, Behavioural Perspective Model variables in the form of Informational and Utilitarian (positive and negative) reinforcement and marketing variables. There is very little which focusses on a hierarchical structured model within the BPM literature. This text will therefore explore the concept of whether any further benefit can be gained through the introduction of such a structure and whether the BPM can facilitate this. Finally, the area of Bayesian estimation has been growing over recent years and this can benefit the field of consumer psychology through the introduction to additional analytical tools which have unique properties given more mainstream analytical techniques such as frequentist measures. This Bayesian inference will be applied to the BPM framework through the development of a Bayesian hierarchical modelling framework.

Chapter 3: Data discussion and category review

3.1 Data Discussion

The data relate to a panel sample of 1,689 consumers/households and 141,592 purchases of four categories within the GB fast moving consumer goods (FMCG) market, as captured by TNS consumer panel. The categories within the data comprise: 1,594 households and 75,563 purchases from the biscuits (sweet and savoury); 895 households, 21,394 purchases from the fruit juice (fruit and vegetable) category; 832 households, 30,906 transactions from the yellow fats (including oils and spreadable) category; and 832 households, 13,729 transactions from the baked beans (including flavours and added ingredients, such as sausages) category. The data accounts for the period of week ending 17 July 2004 to 9 July 2005. The households within the data may purchase any number of items from any number of the four categories. The data are assembled at stock keeping unit (SKU) level, whereby each descriptor contains a string relating to the brand, the number of items within the pack.

3.1.1 Biscuits

Within the biscuits category, the nature of the packets of biscuits relies on the type of biscuit they contain. (e.g. whether they relate to individual based biscuits for example “KitKat” or many smaller biscuits such as digestives). It is conceivable that a packet of digestives may contain many individual biscuits though the packet itself is seen as an individual item, whereas a six pack of KitKat would be seen by the consumer as six individual items grouped into a larger size product offering. Hence it is important to understand how these are classified as it is not consistently coded within the data. It is important to understand this information as the price variable is priced per product, therefore it is imperative to know if the price would relate to the total for the six KitKat biscuits (keeping to the example) or whether it relates to each of the six KitKat biscuits.
In order to achieve this, the 2,783 SKUs relating to the biscuit category are individually analysed and the relevant information is consistently extracted and coded. This information relates to the brand name, the weight and the number of items per pack.

Some records appear to have an extremely low price per SKU (as low as 1 pence per item) and a decision is required as to how these observations are treated. The lowest value biscuits range are classed as supermarket own label or value brands. There is a minimum price of circa 20p per pack. Hence a minimum price of 20p is used as a minimum acceptable price for a packet of biscuits. Any transactions at the SKU level that place a biscuit pack lower than 20p per item are excluded from the study.

In the same manner, there are transactions with a very low price per 100g. Likewise an analysis of the supermarket value range suggests a cut-off point of 15p per 100g is appropriate and hence this is used as a cut off floor for all transactions. This leaves a base sample of 61,087 records to analyse (80.8% of the original biscuits category panel data).

As well as the SKU name, there is a product description field. Table 1 shows the distribution of data within this.

	Count	% Count	Volume	% Volume
BISC CHOC COUNTLINES	17,293	28.3%	5,089,771	28.3%
BISC CHOC FULLY COATED	1,715	2.8%	468,712	2.6%
BISC CHOC SEMI CTD/LATTICED	7,033	11.5%	2,736,751	15.2%
BISC COCONUT	397	0.6%	119,320	0.7%
BISC CREAM/JAM FILLED INC SANDWICH	3,381	5.5%	986,035	5.5%
BISC DIGESTIVES EXC CHOC	526	0.9%	176,500	1.0%
BISC FRUIT FILLED	1,910	3.1%	552,825	3.1%
BISC GINGER	1,124	1.8%	362,650	2.0%
BISC MARSHMALLOW FULLY CHOC CTD	821	1.3%	191,293	1.1%
BISC MARSHMLLS CHOC SEMI/UNCOATED	76	0.1%	17,484	0.1%
BISC SAV CRISPBRD/RICE CAKES	5,332	8.7%	913,185	5.1%
BISC SAV EXTRUDED CRCKRS/WATERBISCS	2,723	4.5%	775,180	4.3%
BISC SAV REMAINING	6,650	10.9%	1,543,272	8.6%
BISC SHORTBREAD	1,295	2.1%	443,455	2.5%
BISC SHORTCAKES	1,295	2.1%	381,622	2.1%
BISC SWEET REM TYPES	6,197	10.1%	2,001,843	11.1%
BISC SWEET/SEMI SWEET ASSORTMNT	913	1.5%	601,675	3.3%
BISC TEA & COFFEE	1,062	1.7%	324,066	1.8%
BISC WAFERS	1,344	2.2%	285,418	1.6%

Table 1: Distribution of data - biscuits

From inspection of Table 1 there are many categories of biscuit that can be confusing to understand and probably not how the consumer would classify purchases. Also, there are some categories with a relatively low number of transactions, which may lead to mathematical sample size issues when analysing. In order to overcome these issues, categories are grouped together to form logical macro categories. Chang (2007, p. 107) suggests a 5 band classification yielding the following groups (Table 2) which is deemed to be a sensible approach and undertaken in this study.

Subcategory	Definition
Chocolate countlines	Individually wrapped chocolate-covered cookie bars which can be sold in multipacks, including Penguin, Club, Breakaway, classic, Kit-Kat, Twix etc., which are marketed and packaged both as confectionary and biscuits.
Plain sweet biscuits	Plain sweet biscuits are uncoated, untopped or unfilled but can be flavoured, for example, coconut or chocolate, including chocolate chips, digestives, sweet assortment, shortbread, shortcakes, wafers, coconut, tea/coffee biscuits and ginger.
Chocolate coated biscuits	Plain sweet biscuits coated partially, topped or completely with chocolate
Filled biscuits	Sweet biscuits which can either be filled, topped or sandwiched between plain biscuits
Non-sweet biscuits	Plain savoury biscuits, savoury crackers and bread-like savoury biscuits. Often flavoured or topped with salt, cheese or other savoury products.

Table 2: Category definition

This grouping of the categories results in the distribution of items shown in Table 3. This grouping is now more identifiable to consumer and have sample sizes that allow more robust statistical analysis to be undertaken.

	Count	% Count	Volume	% Volume
Countlines	17,293	15.2%	5,089,771	14.9%
Chocolate_coat	9,645	8.5%	3,414,240	10.0%
Plain_sweet	14,153	12.4%	4,696,549	13.7%
Filled	5,291	4.6%	1,538,860	4.5%
Non_Sweet	14,705	12.9%	3,231,637	9.4%

Table 3: Regrouped distribution of data - biscuits

From the SKU field, it is possible to identify the number of items per pack. For the reasons stated earlier as to how packs are defined in terms of number of items, a manual process is conducted to allocate the items per pack to each SKU. For example, a packet of say 50 digestive biscuits is seen to be 1 unit whereas a family pack of six KitKats would be deemed to hold 6 items. Therefore, for each SKU, the number of items per pack is extracted manually. The resulting number ranges from as few as 1 biscuit per pack to 48 biscuits per pack. There are also other larger formats such as drums, bags, barrels etc. which do not contain the actual number of items but all imply larger packs. A sensible structure for analysis purposes is required. Hence the biscuit pack sizes are grouped as per Table 4 based on both the distribution of transactions within group and logical groupings. Note that for packets of biscuits which contain many standard biscuits (such as digestive), the figure relates to the number of packets within the SKU, in this case 1. Where biscuits are individually wrapped, biscuits tend to be single serve portions rather than multiple serve. For example, a single packet of six KitKat biscuits will be classed as "6". The item "pack" relates to larger units where the number actual number of biscuits within is not stated on the packaging, e.g. barrels or assortment tubs.

	Count	% Count	Volume	% Volume
1	33743	55.2%	9,354,867	52.1%
2-5	3922	6.4%	1,195,547	6.7%
6-7	6880	11.3%	1,665,044	9.3%
8-11	7349	12.0%	2,056,553	11.4%
12+	6771	11.1%	2,597,211	14.5%
pack	2422	4.0%	1,101,835	6.1%

Table 4: Items in pack distribution - biscuits

This resulting pack distribution is both logical and appropriate for statistical analysis. Most of the category is constructed of the single pack size, though larger packs (i.e. the 12 and the "pack") sizes have a higher volume per transaction as may be expected from larger formats. A similar exploratory analysis is performed on the other categories

3.1.2 Fruit Juice

The initial exercise is to extract the juice type and number of items in the pack from each of the transaction records. As with the biscuit category, all possible SKUs are considered and the two variables are extracted in each case. Table 5 shows the distribution of the number of transactions and distribution of volume (in ml) associated with each of the juice types.

	Juice Type			
	Count	Count %	Volume	Volume %
?				
Apple	2,991	17.4%	6,747,240	19.5%
Blackcurrant	12	0.1%	13,000	0.0%
Breakfast	77	0.4%	110,000	0.3%
Clementine	20	0.1%	28,000	0.1%
Cranberry	44	0.3%	50,000	0.1%
Grape	525	3.1%	813,000	2.4%
Grapefruit	846	4.9%	1,355,750	3.9%
Mango	8	0.0%	6,000	0.0%
Melon	1	0.0%	1,000	0.0%
Mixed	1,439	8.4%	2,193,350	6.3%
Orange	9,652	56.2%	21,162,580	61.2%
Peach	17	0.1%	28,000	0.1%
Pineapple	982	5.7%	1,375,500	4.0%
Prune	55	0.3%	59,000	0.2%
Raspberry	2	0.0%	1,500	0.0%
Strawberry	2	0.0%	600	0.0%
Tomato	368	2.1%	484,994	1.4%
Variety	42	0.2%	64,200	0.2%
Vegetable	54	0.3%	65,472	0.2%
Vitamin	27	0.2%	30,330	0.1%

Table 5: Distribution of data - fruit juice

From inspection of Table 5 it can be seen that some categories account for a small proportion of both transactions and volume, making analysis of the individual categories difficult. Orange and Apple flavours dominate the category, accounting for 73.6% of transactions and 80.7% of volume. There are numerous categories which account for a small percentage of transactions and volume with the notable examples of Melon, Mango, Raspberry and Strawberry all of which have fewer than 10 transactions for the entire period. To accommodate these low counts some flavours are grouped together. First, the juice type variable is grouped in a sensible manner to avoid small categories. Hence the flavour “Other fruit” is created which represents blackcurrant, clementine, cranberry, mange, melon, peach, prune, raspberry and strawberry. Also, the flavour “Variety” is grouped together with the flavour “Mixed” given the similarity of their meaning and the fact they are both small flavours in terms of transactions and volume. The result is shown in Table 6 below which now has fewer but more robust categories whilst maintaining a logical composition.

Juice Type Reduced				
	Count	Count %	Volume	Volume %
Apple	2,991	17.5%	6,747,240	19.5%
Other_fruit	161	0.9%	187,100	0.5%
Breakfast	77	0.4%	110,000	0.3%
Grape	525	3.1%	813,000	2.4%
Grapefruit	846	4.9%	1,355,750	3.9%
Mixed	1,439	8.4%	2,193,350	6.3%
Orange	9,694	56.6%	21,226,780	61.4%
Pineaple	982	5.7%	1,375,500	4.0%
Tomato	368	2.1%	484,994	1.4%
Vegetable	54	0.3%	65,472	0.2%

Table 6: Regrouped distribution of data - fruit juice

The distribution of transactions and volume by pack size is shown in Table 7. The category is dominated by the single serve which accounts for 88.4% of transactions and 81.7% of volume. Therefore, despite its dominance the volume per serve is less than the category average. The four pack has the highest volume per transaction where 4.7% of the category transactions account for 10.5% of the volume, hence this format appears to be a worthy vehicle for category expansion. Other than the 3-pack, all other formats have fewer transactions, hence the variable is grouped to form a variable more suitable for analysis purposes.

Number in Pack				
	Count	Count %	Volume	Volume %
1	15,179	88.4%	28,276,576	81.7%
2	73	0.4%	214,000	0.6%
3	808	4.7%	1,251,150	3.6%
4	740	4.3%	3,639,420	10.5%
5	42	0.2%	67,500	0.2%
6	266	1.5%	949,500	2.7%
9	39	0.2%	76,050	0.2%
10	9	0.1%	18,000	0.1%
12	8	0.0%	84,000	0.2%
27	3	0.0%	16,200	0.0%

Table 7: Items in pack distribution - fruit juice

Given the shape of the distribution it would seem sensible to maintain the single serve as its own serving. The next group consists of 2 through to 5 packs and six packs or larger forms a third grouping. The resulting distribution is shown in Table 8. The larger pack sizes have a larger volume per transaction than the single serve formats, given the volume % represents a larger proportion than the count %.

Number in Pack Grouped				
	Count	Count %	Volume	Volume %
1	15,179	88.4%	28,276,576	81.7%
2-5	1,663	9.7%	5,172,070	15.0%
6+	325	1.9%	1,143,750	3.3%

Table 8: Regrouped items in pack distribution - fruit juice

3.1.3 Yellow Fats

A similar approach is applied to the yellow fats category. Some records are discarded due to the calculation of price per 100g giving an unrealistically low value for some branded items, significantly lower than the supermarket own brand. Therefore, any item with price per 100g below 3.6 pence is discarded. This excludes 129 records. The type of fat is categorised into four types and are shown in Table 9 below.

Fat type				
	Count	Count %	Volume	Volume %
Blended spreads	7,742	31.9%	5,648,250	37.9%
Butter	8,132	33.5%	3,619,014	24.3%
Margarines	4,561	18.8%	3,050,750	20.5%
Low_Reduced	3,810	15.7%	2,578,500	17.3%

Table 9: Distribution of data - yellow fats

Table 10 shows the distribution of number in pack for the yellow fats category. Yellow fats are predominantly sold as single items, accounting for 99.5% of transactions and 99.2% of volume.

Number in Pack				
	Count	Count %	Volume	Volume %
1	24,131	99.5%	14,781,544	99.2%
2	45	0.2%	45,500	0.3%
3	40	0.2%	54,750	0.4%
4	29	0.1%	14,720	0.1%

Table 10: Items in pack distribution - yellow fats

Therefore, all packs with more than one item are combined into a single group which represents two items or more. Larger pack account for a higher average volume per transaction as seen it previous categories. This distribution of the new grouped pack size is shown in Table 11.

Number in Pack Grouped				
	Count	Count %	Volume	Volume %
1	24,131	99.5%	14,781,544	99.2%
2+	114	0.5%	114,970	0.8%

Table 11: Redistributed items in pack distribution - yellow fats

3.1.4 Baked Beans

The fourth category to enjoy this type of analysis is the baked beans category. Sixty two records have no weight recorded against them, hence not possible to calculate total volume or price per 100g and are therefore discarded. The resulting distribution of beans by assortment type is shown in the Table 12 below.

Beans Type				
	Count	% Count	Volume	% Volume
BBQ	47	0.4%	36,339	0.3%
Beans	6,117	56.8%	6,286,269	57.6%
BeansNuggets	17	0.2%	15,595	0.1%
BeansSausage	1,362	12.6%	1,368,244	12.5%
BeansVegiSau	38	0.4%	38,865	0.4%
BeansWedge	14	0.1%	13,355	0.1%
Boston	1	0.0%	1,134	0.0%
Breakfast	199	1.8%	186,891	1.7%
Burger	1	0.0%	220	0.0%
Cheese	17	0.2%	18,950	0.2%
Chilli	16	0.1%	8,765	0.1%
Cone	11	0.1%	10,040	0.1%
Curry	148	1.4%	130,438	1.2%
HealthyBeans	640	5.9%	708,061	6.5%
Mexican	26	0.2%	28,560	0.3%
Omelette	11	0.1%	12,610	0.1%
Organic	4	0.0%	3,750	0.0%
Tomato	2,100	19.5%	2,041,274	18.7%

Table 12: Distribution of data - beans

This category shares a similarity with fruit juice in that it is dominated by a few variants which account for a large proportion of both transactions and volume. Also, there are some variants with very few transactions. These smaller sized categories are grouped together to form the reduced type of “BeansPlus” which relates to Beans with “nuggets”, “sausage”, “vegetarian sausage”, “wedges”, “breakfast”, “burger”, “cheese”, “cones” or “omelette”. Similarly, the category “Flavours” is constructed which represents the flavour variants of “Boston”, “chilli”, “curry” and “Mexican”. Finally, the variant “Healthy” is created to represent the “reduced salt” variety and the “organic” brands. The grouped variables are shown in Table 13.

Beans Type Grouped					
	Count	% Count	Volume	% Volume
Beans	6,117	56.8%	6,286,269	57.6%
BeansPlus	1,670	15.5%	1,664,770	15.3%
Tomato	2,100	19.5%	2,041,274	18.7%
Healthy	644	6.0%	711,811	6.5%
Flavours	238	2.2%	205,236	1.9%

Table 13: Regrouped distribution of data - beans

Table 14 shows the number of items in pack for the Beans category. The category is dominated by two pack sizes, namely the single item and the 4-pack. These account for 97.8% of transactions and 94.3% of the volume of the category.

Number in Pack				
	Count	% Count	Volume	% Volume
1	8,541	79.3%	6,319,330	57.9%
2	5	0.0%	4,150	0.0%
3	10	0.1%	6,000	0.1%
4	1,993	18.5%	3,971,660	36.4%
6	211	2.0%	562,860	5.2%
12	9	0.1%	45,360	0.4%

Table 14: Items in pack distribution - beans

A decision is made to use the two dominant pack size variants as the basis of two encompassing pack sizes, given the dominance of each. Therefore, two groups are created, one which represent packs with 1, 2 or 3 items and a second which represents packs with 4 to 12 items, shown in Table 15. The larger pack sizes account for a larger volume per pack as is seen with past categories.

Number in Pack Grouped				
	Count	% Count	Volume	% Volume
1-3	8,556	79.5%	6,329,480	58.0%
4-12	2,213	20.5%	4,579,880	42.0%

Table 15: Redistributed items in pack distribution - beans

3.2 Conclusion

Now the data have been cleansed and regrouped into meaningful and statistically robust groups, the analysis of the data can begin.

3.3 Category Overview Analysis

In order to better understand each of the category dynamics, an analysis is undertaken for each category in turn. Each analysis will focus on the variables which have been identified within the literature review namely the behavioural economic variables and the Behavioural Perspective Model variables.

3.3.1 Construction of the Informational and Utilitarian Reinforcement Variables

Each product is assigned an Informational and Utilitarian reinforcement value that is the basis of the BPM. The Informational reinforcement may be the brand name or brand differentiation of the product. It may be seen as being akin to the equity associated with the product (Oliveira-Castro *et al.*, 2006). The informational score was derived through a questionnaire related to the awareness of the brand and its perceived quality. The informational score is a computed average based on the two scales and is a scalar variable with a similar scale across all products. The questionnaire was administered by 33 experts, and the results between the 33 were verified. (For more details, see Oliveira-Castro *et al.*, 2006).
The utilitarian reinforcement denotes the functional, practical, and economic attributes of the product. Each SKU within the panel data has a descriptor field which has details about each product. Through this, each SKU was allocated to two levels of Utilitarian reinforcement: level (1) which is a lower utilitarian level and (2), which denotes a higher utilitarian level. The lower level (1) is a base entry level of the product, whereas the higher level (2) denotes a higher quality, more valuable functionality or increased physical attributes. It generally has a higher price point (Chang, 2007). The utilitarian reinforcement is a dichotomous variable denoting the lower or higher level.

3.4 Biscuits

3.4.1 BPM variables

3.4.1.1 SKU count vs. Informational reinforcement

Fig 3 shows the average informational scores per defined stock keeping unit (SKU). The y-axis is the number of SKUs and the x-axis represents the mean informational reinforcement score for the SKU. SKUs with a larger number of variants tend to have a higher informational score. This is demonstrated through higher (*x*, *y*) Cartesian values. There are a large number of supermarket own brand SKUs and this is in fact the dominant SKU in the category. This is the outlier point showing there are circa 14,000 SKUs with an Informational reinforcement value of circa 1.8.

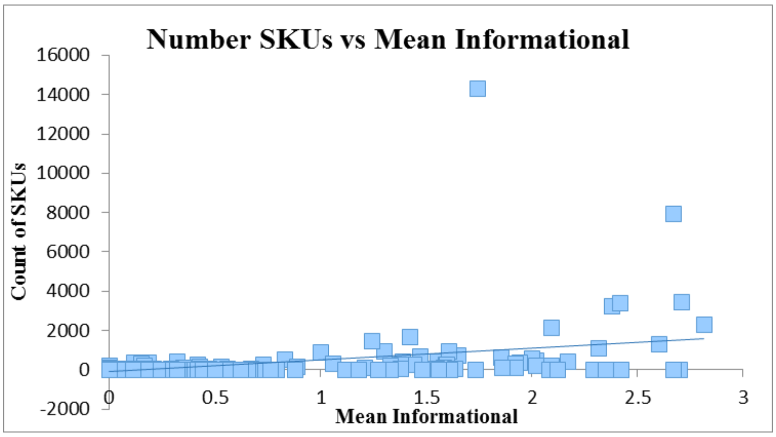

Figure 3: Number SKUs vs Informational reinforcement - biscuits

3.4.1.2 Size of Brand vs Informational Reinforcement

Fig 4 represents top 68% of units sold. The scatterplot shows the relationship between SKU size (number of units sold) and informational score. There is a group of brands with larger number of SKU variants (indicated within the ellipse) which seems to score a higher informational score and smaller SKUs scoring a lower informational score which may indicate that larger informational reinforcement brands tend to have a higher number of selling SKUs (or indicates that larger SKUs will attract a higher informational reinforcement score). Supermarket own brand score is the lowest among the larger brands with a mean score more reflective of the smaller SKUs.

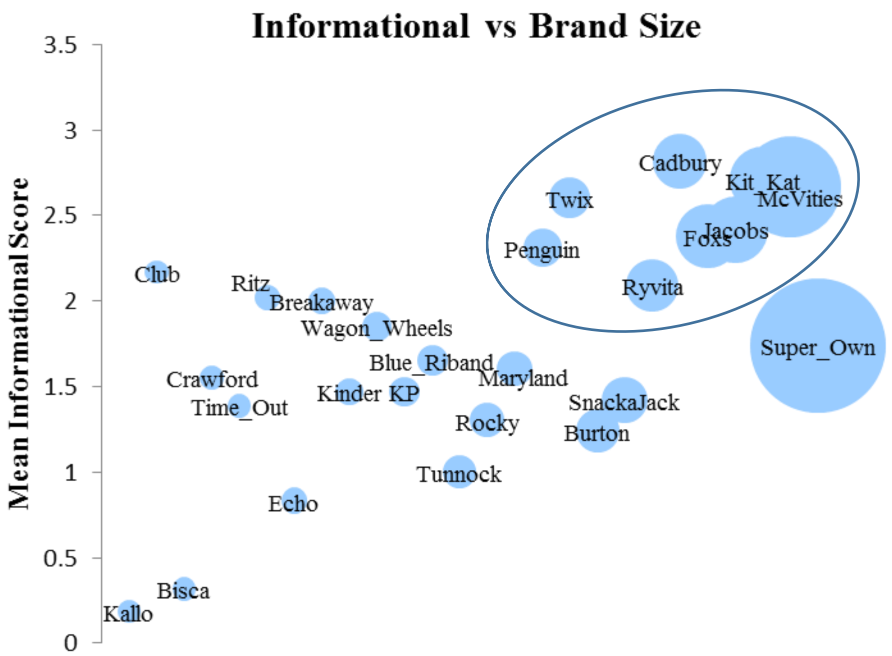

Figure 4: Informational reinforcement vs. brand size - biscuits

3.4.1.3 Biscuit Type and Pack Size vs. Informational Reinforcement

Fig 4 shows the scatterplot of informational reinforcement score and biscuit type and informational score and pack size. The number of units sold is reflective of the bubble size and ordered smallest to largest. There is no apparent relationship between the informational score and either the category size or the biscuit type or the pack size given the biscuit types and pack sizes are all of similar informational reinforcement levels.

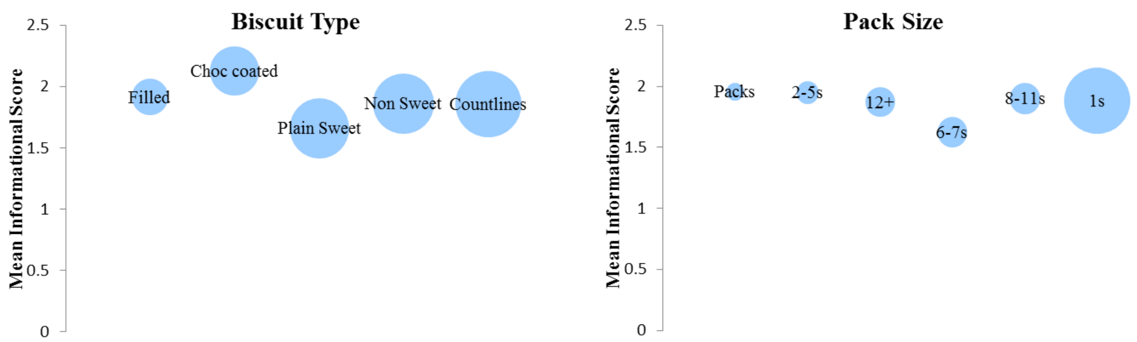

Figure 5: Informational reinforcement vs. type and pack size - biscuits

3.4.1.4 Brand Distribution by Utilitarian Reinforcement

Utilitarian reinforcement variables are dichotomous, representing the lower and higher utilitarian reinforcement levels of the products.
The Venn diagram on the left in Fig 6 shows the distribution of the number of defined brands split by the two utilitarian reinforcement levels. For this category, most brands (59.1%) are located within the lower utilitarian reinforcement level. Some brands can be located within either the lower or higher reinforcement level, depending on the individual SKU within the brand, for example Adams Malted Milk biscuits are coded as utilitarian reinforcement level 1 while Adams Malted Milk with Chocolate is coded as utilitarian reinforcement level 2. This overlap accounts for 21.8% of the defined brands in the category. Utilitarian reinforcement level 2 accounts for 19.0% of the defined brands.

The Venn diagram on the right in Fig 6 shows the volume the SKUs. When considering volume, the largest part of the Venn diagram is the intersection of the two utilitarian reinforcement levels. These 21.8% of brands are accounting for 66.4% of the volume of the defined brand category. This may suggest the variants within the SKU are providing offerings to consumers which purchase within both the utilitarian reinforcement levels. Alternatively, it may suggest that larger brands are able to offer variants which appeal to both utilitarian reinforcement levels. This seems logical as brands which have SKUs representing both utilitarian groups are large enough in volume to sustain more than one brand variant. This may be especially true in a category such as biscuits where additional toppings or flavours could be the difference between a variant being associated with the lower or higher Utilitarian reinforcement levels.

Bottom right of Fig 6 shows the split of the supermarket own brand across the two utilitarian reinforcement levels. There is a relatively even split between the lower and higher utilitarian reinforcement levels which demonstrates the category is driven by the type of biscuit rather than just the branded nature of the biscuit. This implies when it comes to the biscuit category, the own labelled supermarket products are offering a diversity of utilitarian reinforcement attributes.

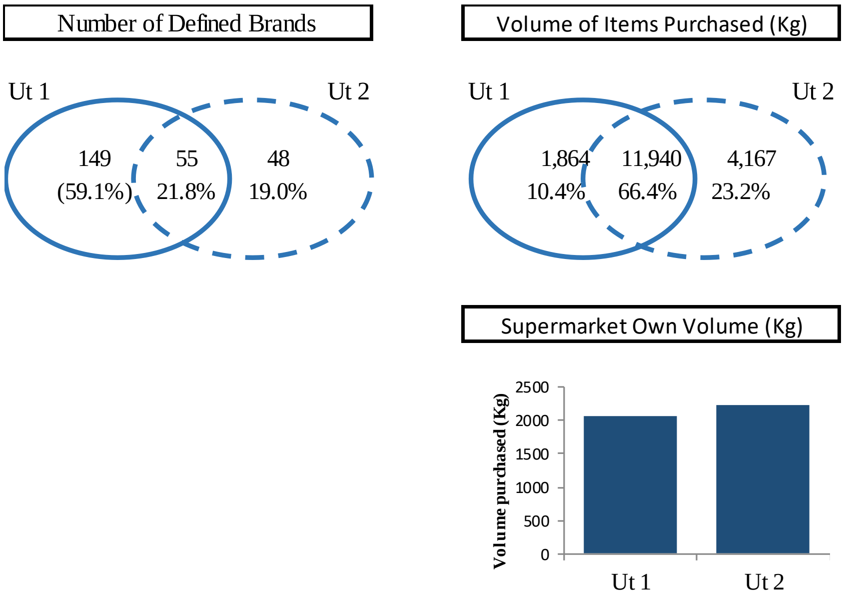

Figure 6: Utilitarian reinforcement analysis - biscuits

3.4.2 Volume Analysis

Fig 7 shows the time series line of volume of the biscuit category split by the utilitarian reinforcement variable. As volume increases during the build up to the Christmas period, it is the lower utilitarian group which increases market share. This seems to suggest consumers are purchasing more utilitarian products for the increased consumption period around the Christmas holidays. The category volume significantly falls at the end of the year. In some part this is due to a shorter commercial week during the Christmas period (fewer shopping days), though this may not be limited to this given the significant drop in sales with only one less shopping day. The volume level returns to near average levels the following week.

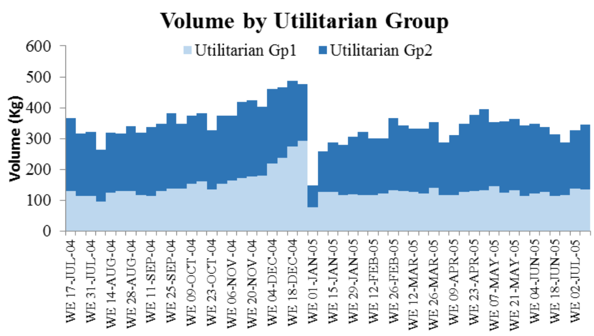

Figure 7: Volume by utilitarian reinforcement group - biscuits

Fig 8 shows the split of volume by biscuit type over the 52 week period. The left hand chart is the seasonal pattern and the right hand chart shows how the share of volume varies across biscuit type by week. The share across type is fairly consistent, despite actual fluctuations in weekly volume (especially the end of year drop). There are some seasonal increases for non-sweet biscuits around the Christmas holiday period. Also, a decrease in the countlines is observed during this period.

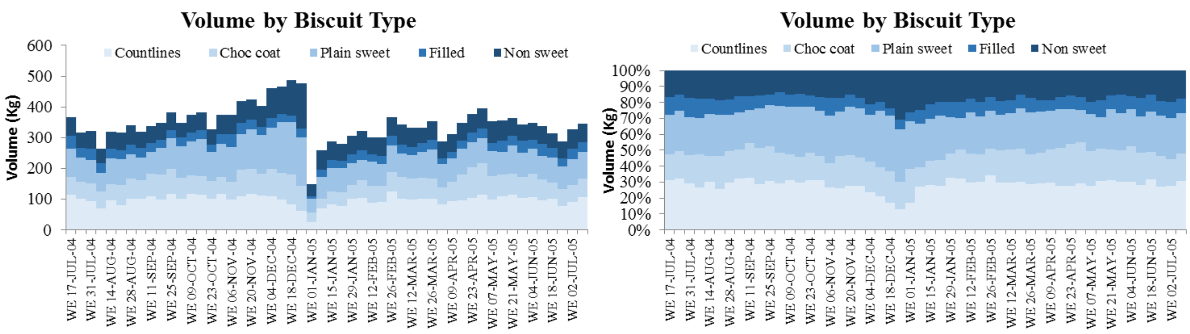

Figure 8: Volume by variety - biscuits

A similar visualization is undertaken for the number of items per pack shown in Fig 9. The single item trend is fairly consistent in terms of weekly share. There is a very seasonal factor apparent with the "pack" format where volume is predominantly driven by the Christmas holiday period (and its run up). This impacts the share of other larger pack sizes.

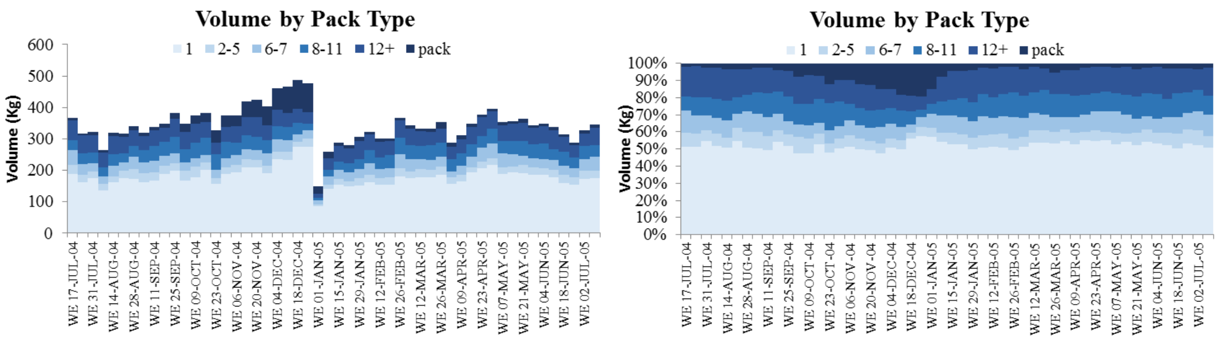

Figure 9: Volume by pack type - biscuits

3.4.3 Average Price

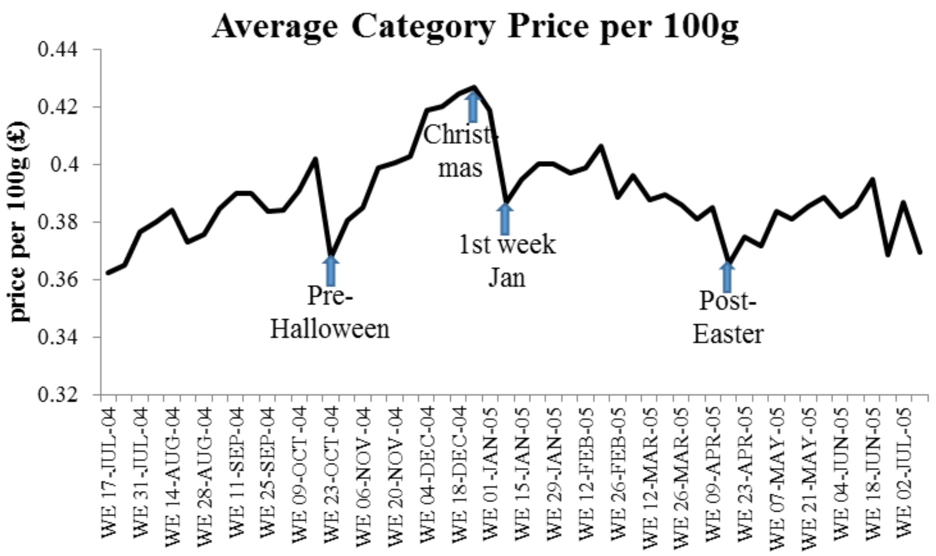

Figure 10: Average price - biscuits

Fig 10 shows average volume weighted price per 100g for the entire category. Average price is shown mathematically in equation 2

$$\text{Avg Price} = \frac{\sum_{i=1}^{n} \text{Volume} \times \text{Price}}{\sum_{i=1}^{n} \text{Volume}} \qquad \text{for brand } i = 1,\ 2,\ \ldots,\ n$$

Equation 2: average price calculation

Some seasonal points have been highlighted which may reflect the relationship between demand and price. It would seem that traditional biscuit peaks such as Christmas has a relatively higher price whereas more traditional lower seasonal periods have a lower average price. The increase in price towards and during the Christmas period may be reflective of the change in number in pack format, where the higher priced "pack" format is more dominant within this period. Also, during this period, it has been observed a decrease in share of the countlines biscuits range over the Christmas period. If this is the case it would suggest the type of biscuit and also the nature of the pack size is contributing to changes in average price and hence should be considered as part of the marketing mix variables.

The price elasticity of demand can be obtained using equation 3. The coefficients obtained are compatible with economic theory and consistent over time (Oliveira-Castro *et al.*, 2006).

$$LogQ_i = \alpha + \beta Log\mathrm{Price}_{\mathrm{i}} + \varepsilon_{\mathrm{i}}$$

Equation 3: Elasticity of demand

Recent studies have indicated the biscuit category data have a negative elasticity of demand and this value is close to –0.5 (Chang, 2007; Oliveira-Castro *et al.*, 2006). These coefficients being less than unitary value demonstrates the category is inelastic, which is consistent with food products in general (Driel *et al.*, 1997).

This analysis relates to the average price movement per 100g as this can be used regardless of the number of items within a pack and the physical size of those items. The natural logarithm is the change in price. Given the volume is also logged, the price elasticity of demand will be the coefficient within the regression equation 3.
The nature of the average price is also worth further comment. The lack of promotional calendar information means the price elasticity will be an average price elasticity, which will be the result of possible regular (long term) price changes, promotional (short term) price discounts and also changes in category mix, either be it in biscuit type or number of items in pack (especially around the Christmas holiday period discussed earlier).

3.4.4 The Double Jeopardy (DJ) Effect

Past studies have shown the DJ effects in terms of category share. Fig 11 shows the pattern of volume for the top 20 brands by largest volume in the category charted against penetration and frequency. The pattern of DJ is observed whereby the larger volume brands have a higher penetration and a higher frequency. Therefore, larger volume brands have a higher level of penetration and a higher frequency of purchase.

The DJ studies speak only of large brands which are substitutable. Ehrenberg *et al.,* (1990) state further that these are substitutable within a blind taste, however this this does not seem to be the case given tastes will be different. The DJ effect shows that larger brands benefit from higher penetration and increased frequency compared to smaller brands. Fig 11 shows the positive relationship between volume and both penetration and frequency, indicating the DJ effect is prevalent in the category. The DJ effect is overarching and while a powerful way of estimating share, delivers no causal effect to the marketing practitioner other than its law like nature to predicting share (Ehrenberg et al, 1990).

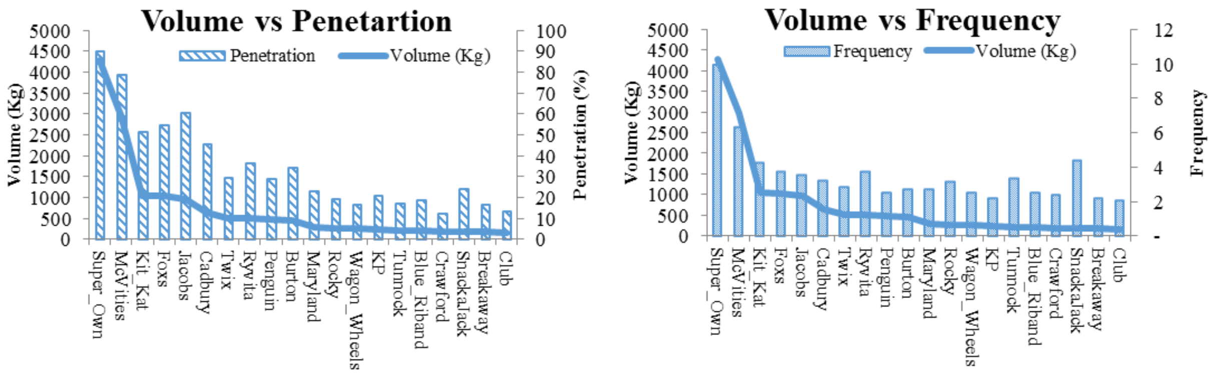

Figure 11: Double jeopardy effect - biscuits

3.5 Fruit Juice

A similar analysis is offered for the fruit juice category.

3.5.1 BPM Variables

3.5.1.1 Informational Reinforcement

Fig 12 shows the number of SKUs pertaining to the fruit juice category plotted by their informational scores. Supermarket own brand has by far the most number of SKUs, with circa 18000 SKUs within this definition. Fruit juice is therefore very dependent on the supermarket own branded products, though both are located around the centre of the Informational reinforcement distribution. In contrast to the biscuit category. There is no correlation between the Informational reinforcement score and the number of SKUs associated with the particular score.
The brands within the category have informational scores ranging from 0.7 to 2.2 with supermarket own brands' value of circa 1.4. Hence the informational reinforcement gained from supermarket own brands is lower than that of the biscuit category.

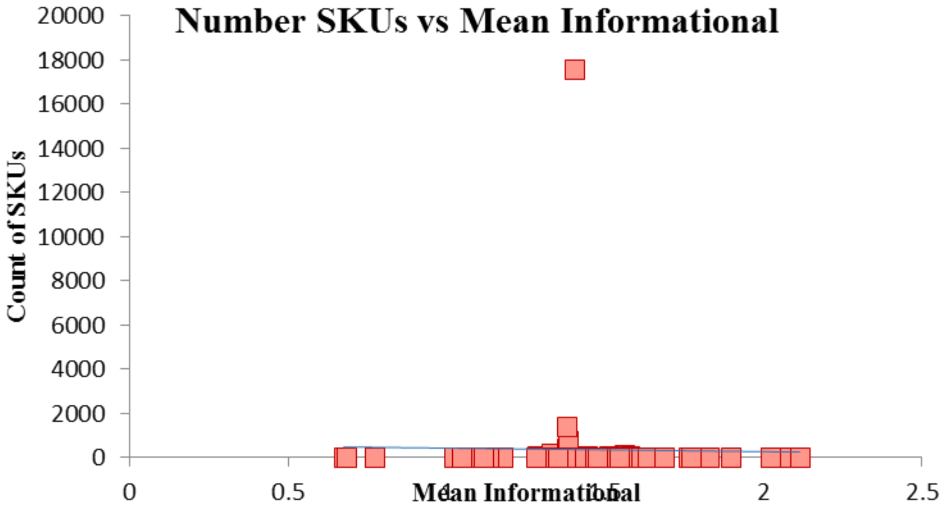

Figure 12: Number SKUs vs Informational reinforcement - fruit juice

3.5.1.2 Size of Brand vs Informational Reinforcement

Fig 13 below shows the brands are in rank order of size (from left to right) and the bubble indicates the relative volume of the brand. The y-axis represents the informational score of the brand. There is a less variation between the size of the brand franchise and the informational score than seen with the biscuit category. Supermarket own brand is very large in terms of its variants, though the mean informational reinforcement score is much in line with other brands, unlike what has been seen in the biscuit category. This may suggest less of a demand for an interaction term of informational reinforcement vs supermarket own indicator. However, note this is the average informational score for the *brand* not the *SKU* and more variance may be observed when the data is modelled at the disaggregate SKU level.

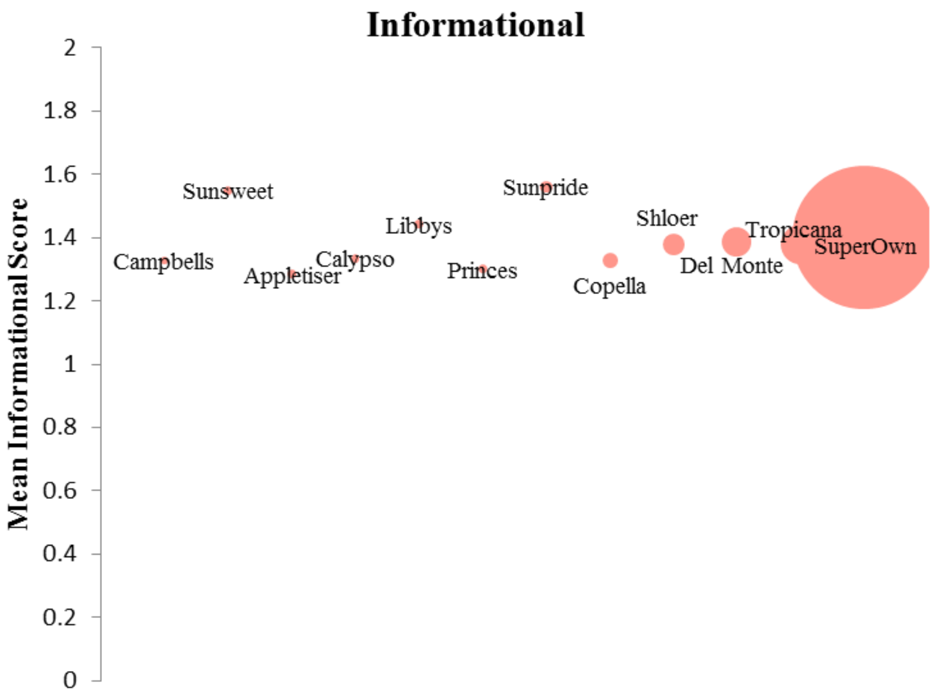

Figure 13: Informational reinforcement vs. brand size - fruit juice

3.5.1.3 Informational Reinforcement by Characteristics

Fig 14 below shows the informational score split by the type of fruit juice and the pack size (as discussed earlier). There is more variance between informational reinforcement score between juice types than was observed for the biscuit types.

In terms of pack size, there is some evidence (though based on three observations) of a decreasing informational reinforcement score as number of items in the pack decreases. However, with such a small sample there is no way of understanding if this is an actual trend or just there through chance.

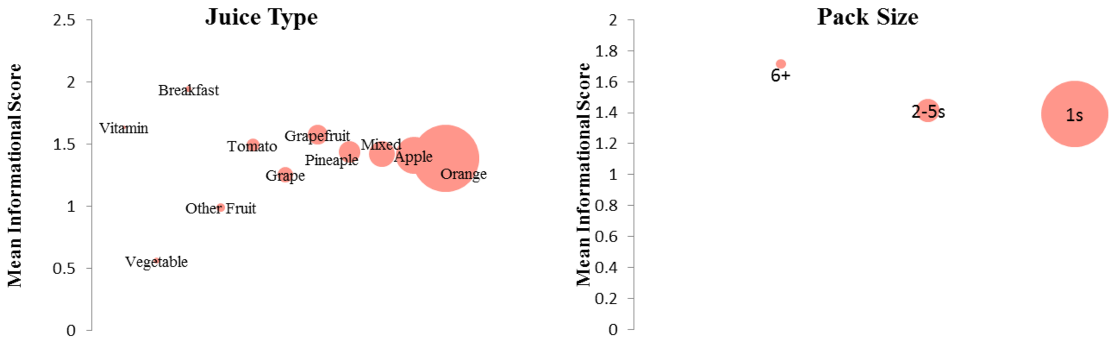

Figure 14: Informational reinforcement vs. type and pack size - fruit juice

3.5.1.4 Utilitarian Reinforcement

Fig 15 shows the graphical analysis of the data by the utilitarian reinforcement variable. The top left Venn diagram entitled “Number of Defined Brands” shows the number of branded SKU variants in each utilitarian reinforcement level. There are only 2 SKUs which are purely of higher utilitarian reinforcement (3.4% of total SKUs). The branded SKUs which are purely lower utilitarian reinforcement level account for 50% of SKUs. However, these single-utilitarian SKUs account for a very small percentage of the volume of the branded SKUs (1.7% and 0.2% of volume for the lower and higher utilitarian reinforcement levels respectively, with the remaining 46.6% of branded SKUs accounting for 98.1% of volume of the category).

The bar chart within Fig 15 shows the vast amount of volume of the category appertaining to supermarket own brands are of the lower utilitarian reinforcement categorization.

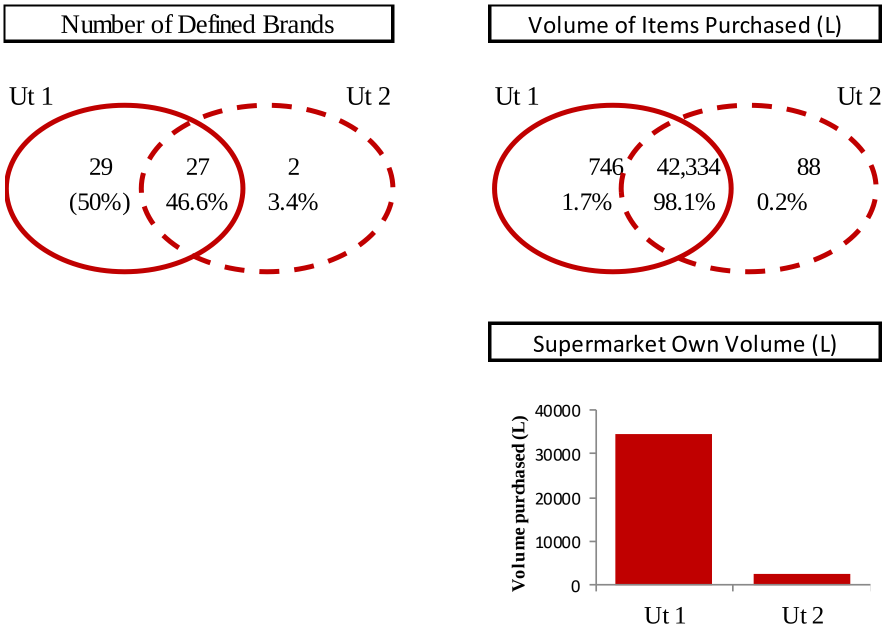

Figure 15: Utilitarian reinforcement analysis - fruit juice

3.5.2 Volume Analysis

The fruit juice category is dominated by the lower utilitarian reinforcement level with little notable variation through the 52 week seasonality pattern shown in fig 16. The last week in the year shows a significant decrease in weekly volume as observed with the biscuit category. Again, note this week has fewer shopping days with the Christmas and New Year holiday periods. There is less evidence of pantry loading prior to the Christmas holiday and the Christmas holiday peak itself is more in line with other weeks through the year.

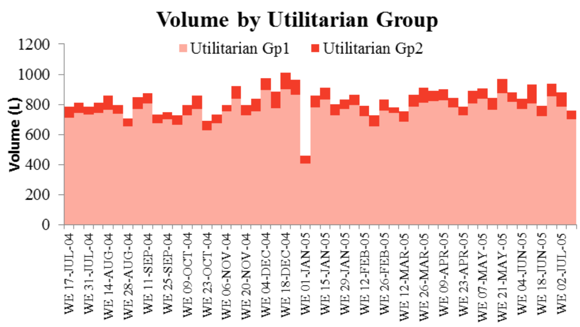

Figure 16: Volume by utilitarian reinforcement group - fruit juice

Fig 17 shows the volume decomposed into defined juice type both on a column chart and on a 100% column chart to visualize category mix change over the year. The chart shows the dominance of the Orange flavoured juice and the Apple as secondary. The 100% stacked bar chart shows no real weekly change in mix through the year. The Christmas period is indistinguishable in terms of juice type share.

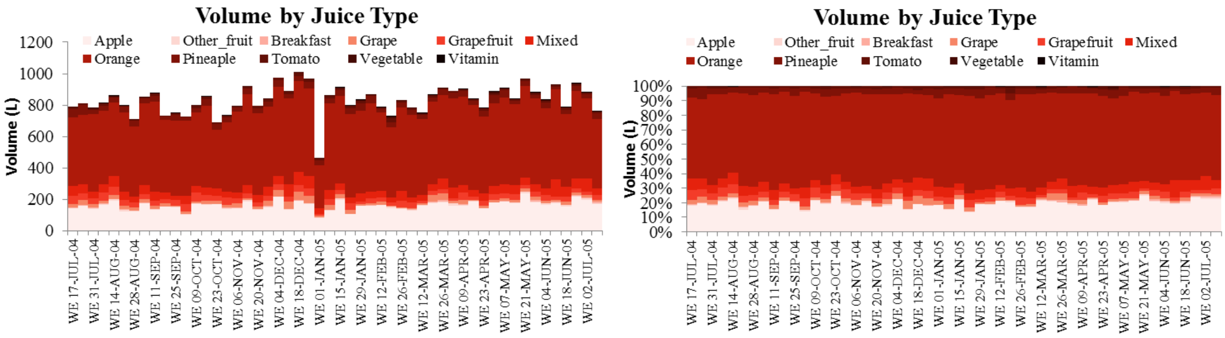

Figure 17: Volume by variety - fruit juice

Figure 18 is the same representation but split by the pack type. The single units dominate the category with no real indication of a change in the pack size mix over the 52 week period (as depicted in the 100% stacked column chart).

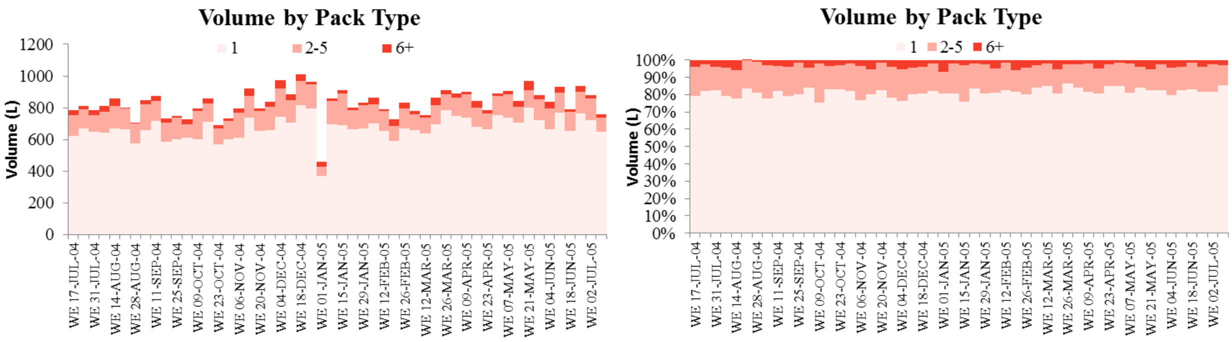

Figure 18: Volume by pack type - fruit juice

3.5.3 Price Analysis

Fig 19 shows the volume weighted average price for the category. The nature of the peaks and troughs of the time series suggests there are no obvious long term price changes in the category and it seems the average price may be driven by seasonal price points and promotional marketing mix mechanics. Given the consistency of the SKU product characteristic mix, it would suggest this is impacting less on the changes in average price versus the biscuits category.

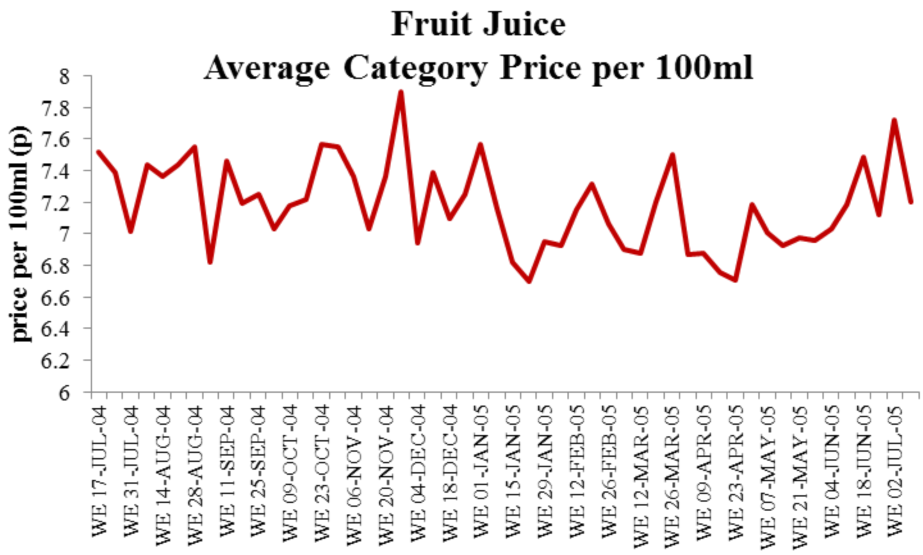

Figure 19: Average price - fruit juice

3.5.4 Double Jeopardy Variables

Fig 20 is a graphical representation of the double jeopardy phenomenon for the top 20 brands (by volume) within the category. The left hand chart shows that smaller volume brands have fewer users (penetration) whereby the right hand chart shows that smaller brands are more likely to have less loyalty (i.e. lower frequency of purchase) than larger brands. This demonstrates the DJ effect is prevalent within the category.

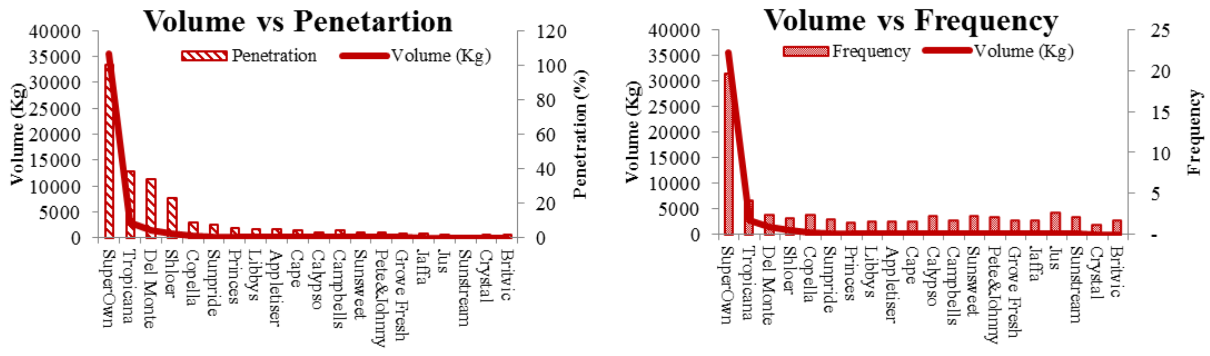

Figure 20: Double jeopardy effect - fruit juice

3.6 Yellow Fats

3.6.1 BPM Variables

3.6.1.1 Informational Reinforcement Analysis

Fig 21 shows the number of brand SKU variants within the yellow fats category and their informational reinforcement score. The y-axis is the number of SKUs associated with the brand and the x-axis shows the brand's average informational reinforcement score across those SKUs. The predominant outlier is that of supermarket own brand, a similar pattern to that observed within the biscuits and fruit juice category. However, the mean informational reinforcement score for these supermarket own brands is relatively lower than for either of the previous categories. The branded items' mean Informational reinforcement levels are larger for the brands with more SKU variants, similar to the biscuit category.

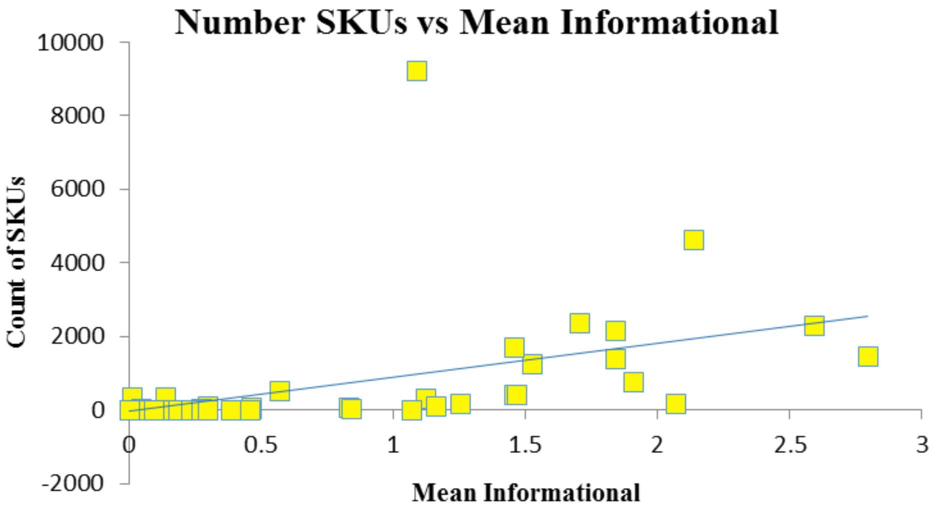

Figure 21: Number SKUs vs Informational reinforcement - yellow fats

3.6.1.2 Size of Brand vs Informational Reinforcement

Fig 22 shows the informational reinforcement score of the top 15 brands. The relationship between the informational reinforcement score and the size of the branded items appears to be positive, with larger brands being associated with higher informational reinforcement. As with the biscuit category the supermarket own brand is the exception with the informational reinforcement being lower than the branded products, given its relative size. This suggests an interaction term between supermarket own brand and the informational reinforcement may be useful.

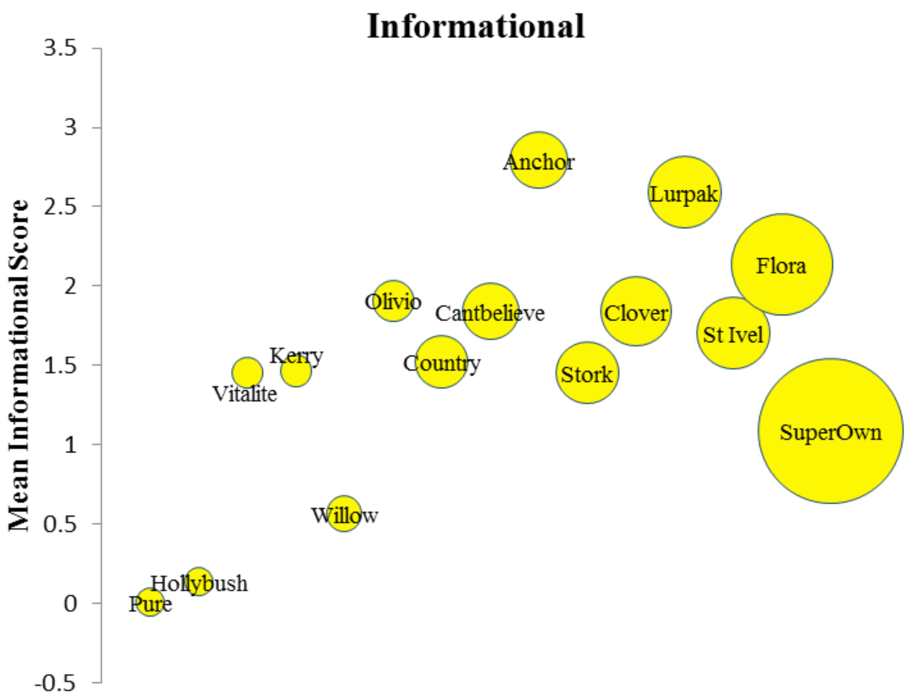

Figure 22: Informational reinforcement vs. brand size - yellow fats

3.6.1.3 Informational Reinforcement by Characteristics

Fig 23 shows the informational reinforcement score split by type and pack size. There is little variation in informational reinforcement by fat type, as seen with the biscuits category. In terms of the number of items in pack, most volume for this category is purchased in single pack sizes and little can be concluded at this stage.

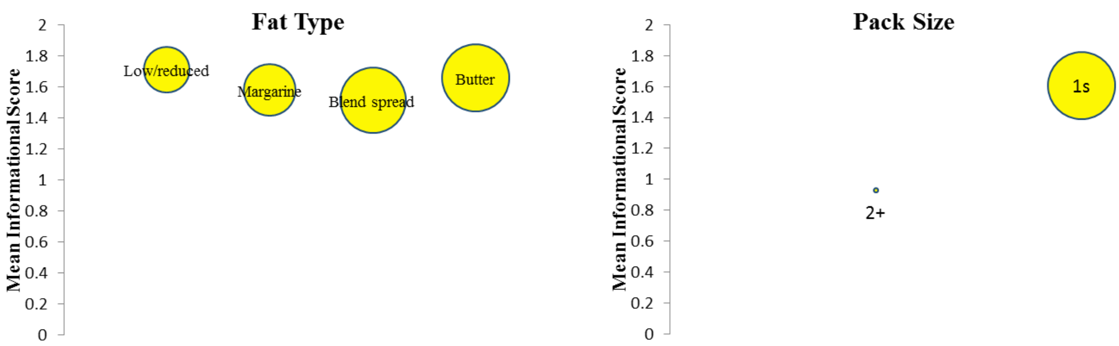

Figure 23: Informational reinforcement vs. type and pack size - yellow fats

3.6.1.4 Utilitarian Reinforcement Analysis

Fig 24 shows the distribution of the category branded items amongst Utilitarian reinforcement levels. Many of the brands (64.6%) are of a lower utilitarian reinforcement level, 14.6% of brands are solely higher utilitarian reinforcement level and 20.8% of brands have SKUs which span both utilitarian reinforcement levels. The volume picture is somewhat different, showing most category volume (60.6%) is accounted for by brands which offer both lower and higher utilitarian reinforcement and only 2.2% of volume is sold through solely higher utilitarian SKUs, while 37.2% of volume is accounted for by the lower utilitarian reinforcement level. The larger volume based intersection of the utilitarian reinforcement levels is a similar picture to other categories within this study thus far.
The bottom right element of Fig 24 shows the volume sold through supermarket own brands is predominantly lower utilitarian reinforcement as seen with the fruit juice category, though different to the biscuit category which was more equally distributed between both utilitarian reinforcement levels.

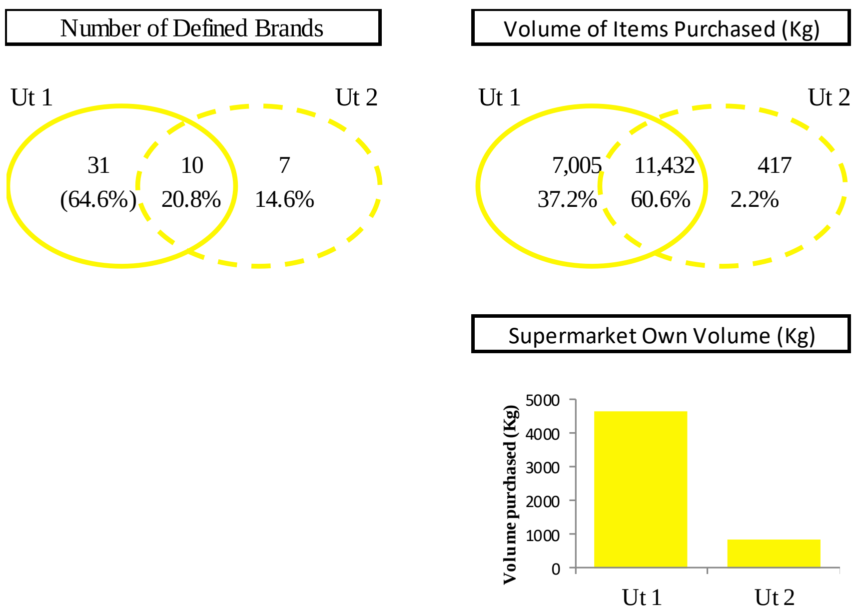

Figure 24: Utilitarian reinforcement analysis - yellow fats

3.6.2 Volume Analysis

Inspection of Fig 25 shows the yellow fats category volume is predominantly associated with the lower utilitarian reinforcement level. There is a seasonal growth at the Christmas holiday period though not significantly higher than other weeks of the year. The spike towards the end of March coincides with Easter weekend.

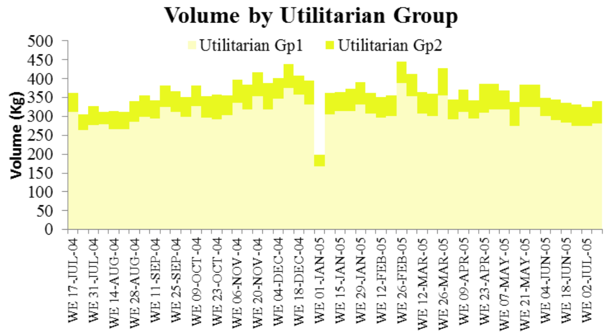

Figure 25: Volume by utilitarian reinforcement group - yellow fats

Fig 26 shows the volume decomposed by yellow fat type. Blended spreads is the largest sub category though does not dominate in the same way as orange juice dominates the fruit juice category. There is a seasonal low for the last full week of December (WE Dec 29th).

The product mix is similar across the 52 week period with some indication of the Low fat/Reduced category having a larger share within the January window, shown by the red arrow. However, this share level is achieved in other parts of the year.

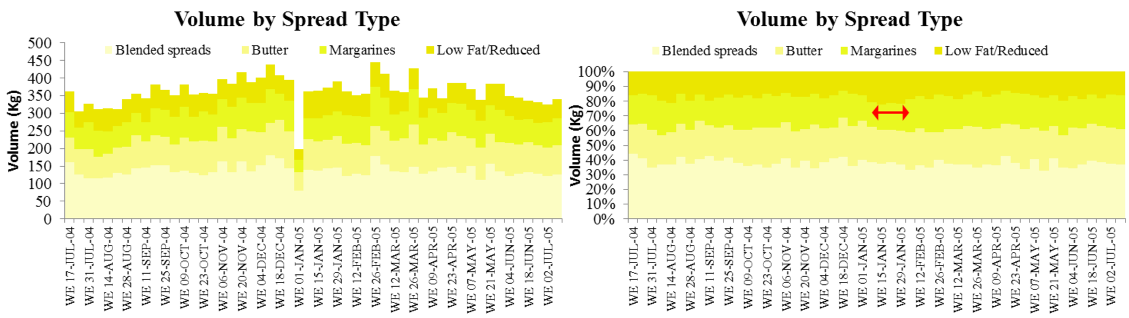

Figure 26: Volume by variety - yellow fats

Fig 27 shows the dominance of the 1 pack purchase for this category. At its weekly peak, the 2+ items in a pack account for 3% of the category share. However, the category is almost entirely dominated by single item products.

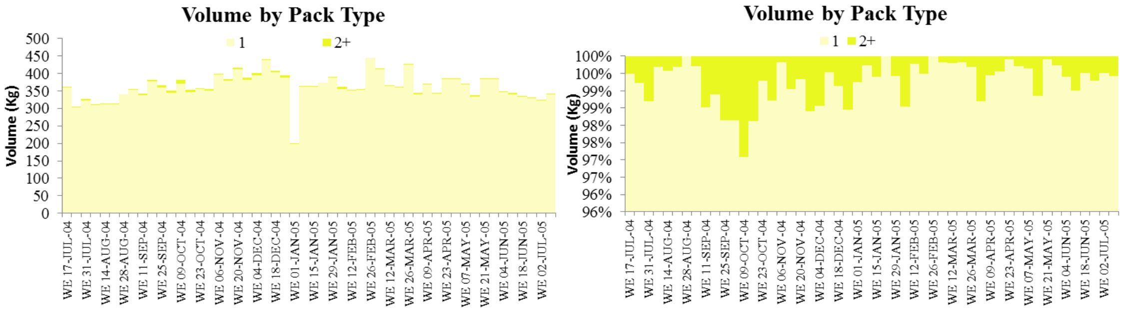

Figure 27: Volume by pack type - yellow fats

3.6.3 Price Analysis

Fig 28 shows the volume weighted average price of the category. There does not seem to be an obvious pattern in the average price and is presumably driven by seasonal and promotional periods. The difference between the lowest and highest average price is very small at approximately 3 pence per 100g.

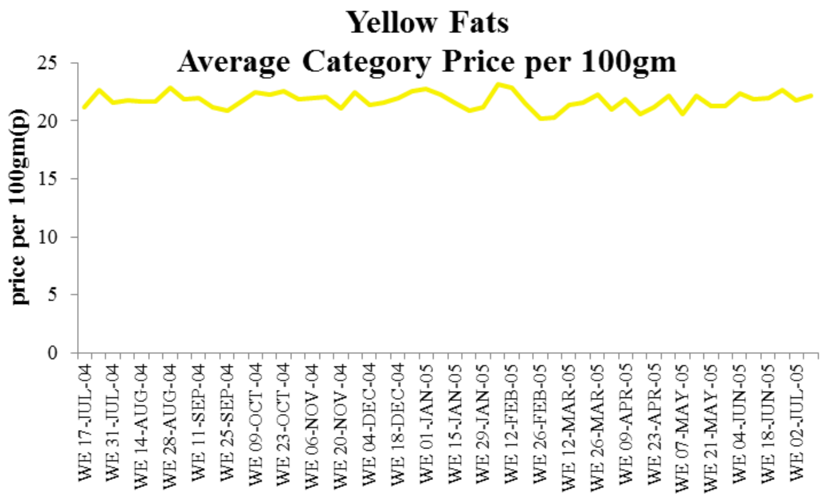

Figure 28: Average price - yellow fats

3.6.4 Double Jeopardy Variables

Fig 29 below shows the double jeopardy variables. As seen with the previous categories, there is a positive relationship between the volume of the brands and their penetration and loyalty.

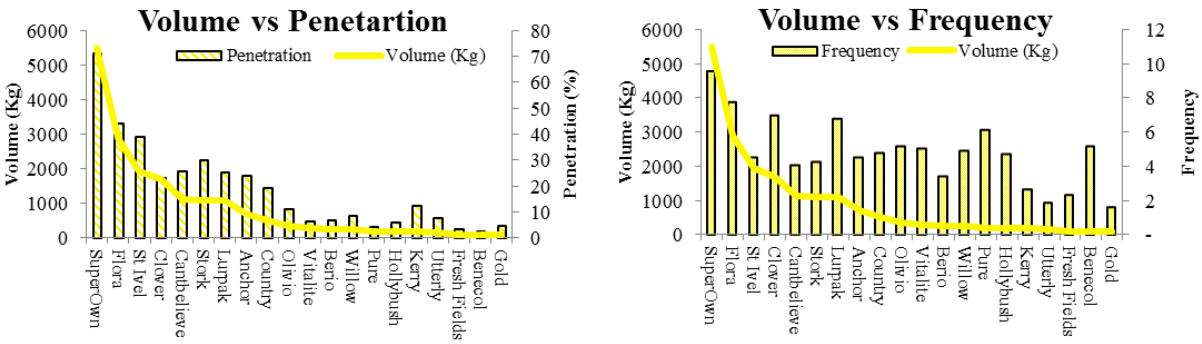

Figure 29: Double jeopardy effect - yellow fats

3.7 Baked Beans

3.7.1 BPM Variables

3.7.1.1 Informational Reinforcement Analysis

Fig 39 shows the number of SKUs per brand of baked beans along the y-axis. There are a large number of supermarket own brand SKUs. Unlike the other categories there is a single dominant branded family of SKUs, namely Heinz. The x-axis is the average informational reinforcement score for each brand group. The supermarket own brand’s Informational reinforcement has an average score of circa 1, similar to the yellow fats category. There are too few brands to hypothesise any trend across the category.

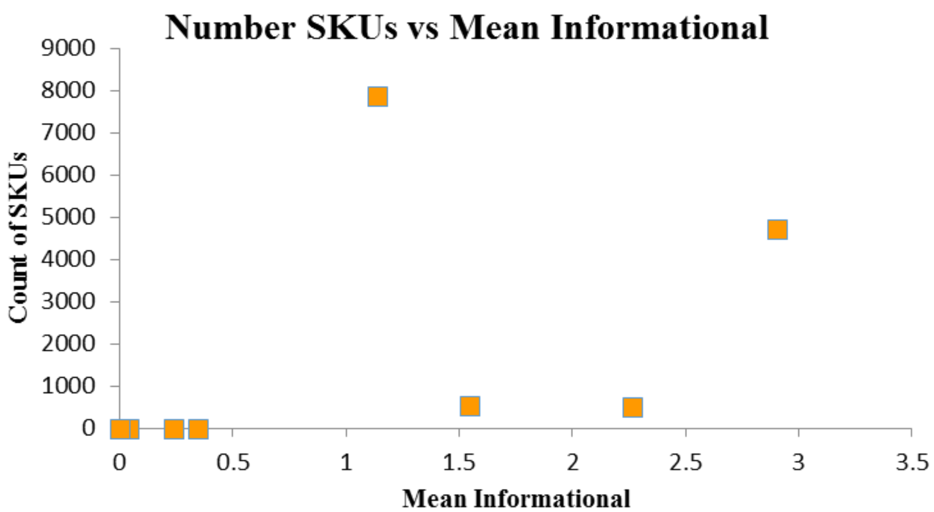

Figure 30: Number SKUs vs Informational reinforcement - beans

3.7.1.2 Size of Brand vs Informational Reinforcement

Fig 31 shows the brands in rank volume order from left to right and the bubbles represent their relative volume size. The y-axis shows the brand's average informational reinforcement score. As with the biscuits and yellow fats category, the larger brands have a higher average informational score. The exception (as with the same two categories) is the supermarket own brand which has a lower average informational reinforcement score than would be expected, given the brand's volume size.

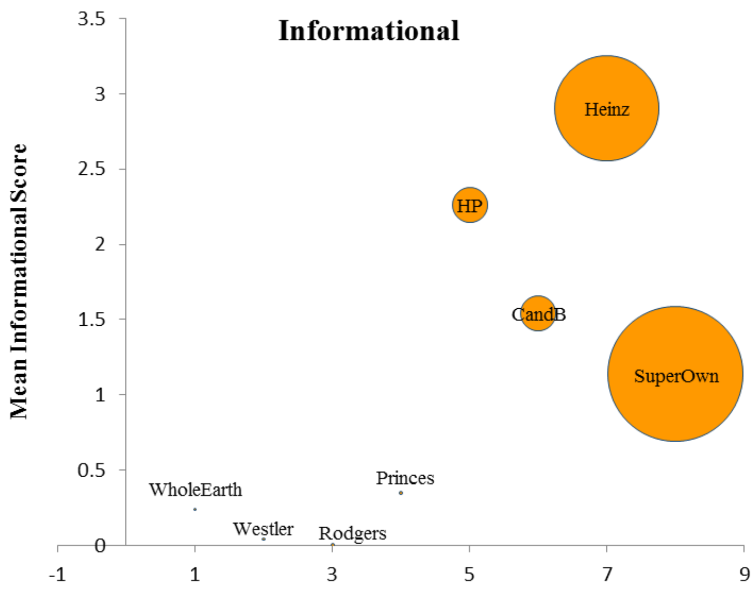

Figure 31: Informational reinforcement vs. brand size - beans

3.7.1.3 Informational Reinforcement by Characteristics

Fig 32 shows the informational reinforcement score by beans type and also pack size. In terms of beans type there is no relationship between the average informational reinforcement score and the beans type. There does seem to be a difference in informational reinforcement between the smaller and larger packs though there are only two data points in the analysis.

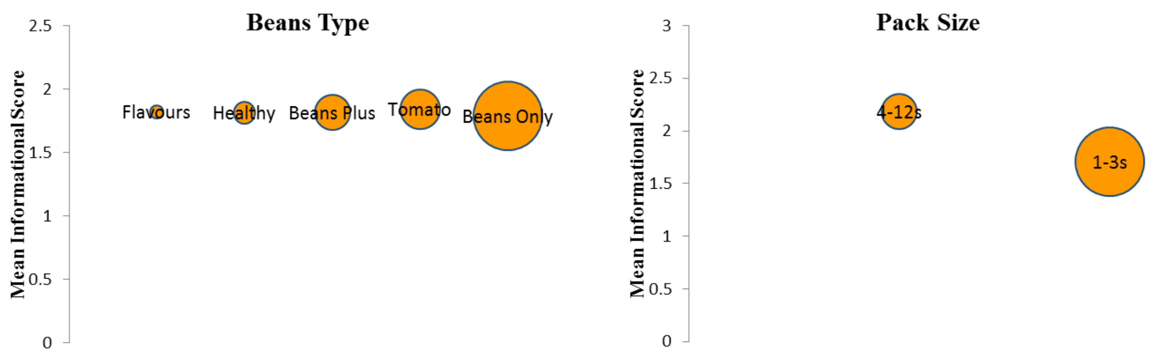

Figure 32: Informational reinforcement vs. type and pack size - beans

3.7.1.4 Utilitarian Reinforcement Analysis

Fig 33 shows the graphical analysis of the utilitarian informational variable. There are only eight branded variants within this category, two within each of the upper and lower utilitarian reinforcement levels and four which have SKUs straddled over both utilitarian levels. Virtually the entire volume are branded items where the SKUs straddle both levels (99.9% of volume) meaning this category effectively has no volume from brands which are entirely within just one of the utilitarian levels.

In terms of supermarket own brand volume, the majority of volume are amongst SKUs within the lower utilitarian reinforcement level, though more evenly distributed than the fruit juice and yellow fats category though less evenly distributed as the biscuits category.

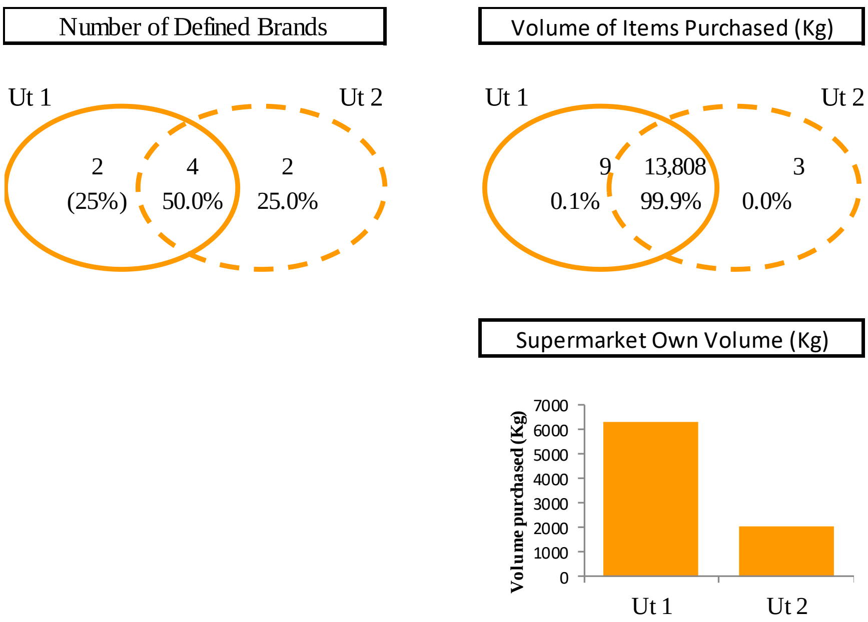

Figure 33: Utilitarian reinforcement analysis - beans

3.7.2 Volume Analysis

A seen in other categories, there is a volume decrease around the Christmas holiday period, shown in Fig 34. The largest week is the first whole week of January, possibly with consumers reverting to more basic consumption following the festive period. The category volume is dominated by the lower utilitarian group and the share across the weeks between the two utilitarian levels is relatively consistent.

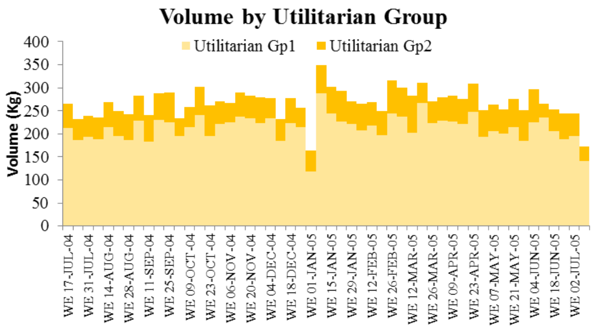

Figure 34: Volume by utilitarian reinforcement group - beans

Fig 35 shows the category's main variant is the beans only format which accounts for circa 50% of the category in terms of volume. The 100% stacked bar chart shows little change in the dynamics of the category across time in terms of the beans type purchased. The larger January volume week does not appear to be significantly different from other weeks in terms of product mix, at least not visually.

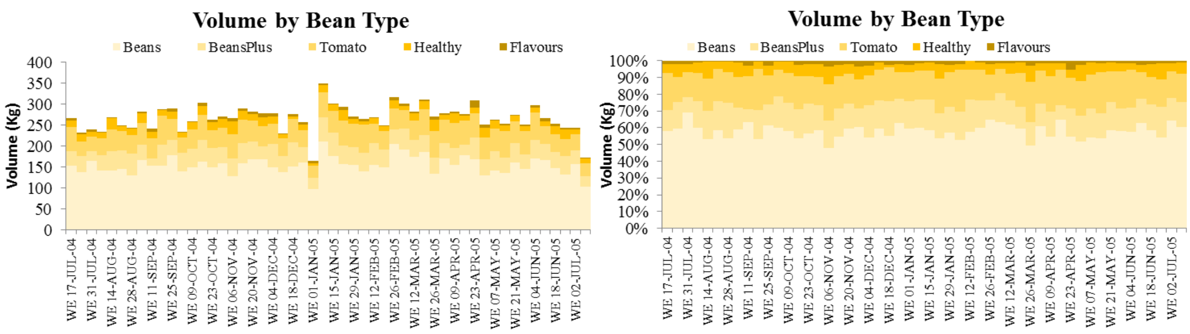

Figure 35: Volume by variety - beans

Fig 36 shows the time series of volume sales split by pack type. There is little change through the year in terms of share between the pack types. The smaller format items per pack is consistently larger in terms of volume share of the category.

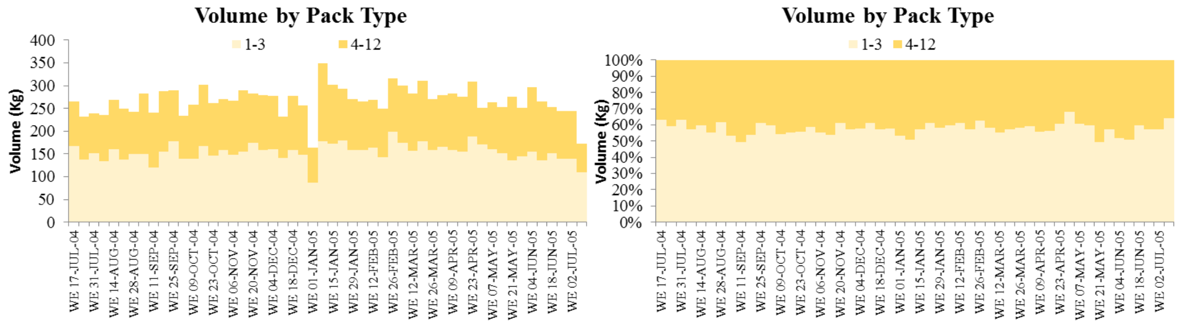

Figure 36: Volume by pack type - beans

3.7.3 Price analysis

Fig 37 shows the volume weighted average price of the category. There has been an increase in price over the 52 week period. The spiked nature of the series is indicative of a promotional category with high and low price points.

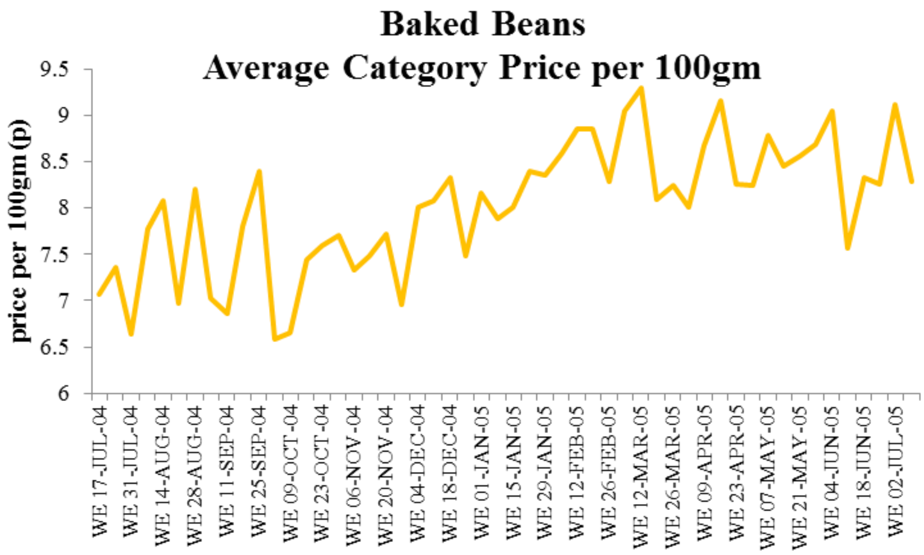

Figure 37: Average price - beans

3.7.4 Double Jeopardy

Fig 38 shows a graphical representation of the volume charted against the penetration and loyalty of the brand families of the categories. As with previous categories, there is a positive relationship between both penetration and loyalty versus the volume of the category, underlining the double jeopardy effect as noted by, for example, Ehrenberg et al (1990).

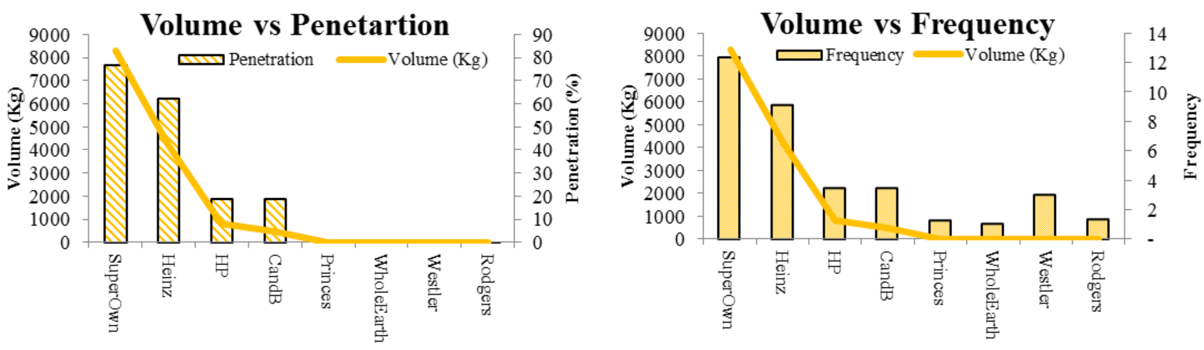

Figure 38: Double jeopardy effect - beans

The study continues by examining the focal variables in turn and assessing their relationship to the volume of each category in turn. This is to determine any relationships existing in the data which may contributing to consumer behaviour. The variables explore the behavioural economics as well as the Behavioural Perspective Model theoretical framework.

3.8 Hierarchical Data Structure

The data represent multiple purchases within a 52 week period amongst 1,689 households. One solution would be to build 1,689 separate models, one for each household. However, the production of such a granular level model has limitations: the process is unrealistically time demanding and often results can return coefficients with incorrect sign and/or unreasonable magnitude (Montgomery and Rossi, 1999).

The data is structured as per Table 16 which shows the number of purchases, number of households and average and median statistics for purchases per household for the total 52 week period. Each category is positively skewed.

	Biscuits	Fruit Juice	Yellow Fats	Beans
Number Purchases	61,081	21,349	30,748	13,660
Number Households	1,592	895	1,354	831
Avg Purchases per Household	38.4	23.9	22.7	16.4
Median Purchases per Houehold	27	17	19	13

Table 16: Distribution of category purchases and purchasers

Many consumer studies are built on aggregated data that can be an issue since the theory is built upon individual behaviour (Kagel *et al.*, 1995). To that end, Oliveira-Castro *et al.* (2006) carried out work that looked at individual consumer elasticity rather than aggregated elasticity, again, within the BPM framework. Oliveira-Castro *et al.* (2006) built individual models for 80 households by using data from the same Fast Moving Consumer Goods product categories comparing individual and household demand. They found that a general assumption of the similar household trends across inter-consumer and intra-consumer could not be made.

Whilst this is interesting, it is challenging for the market researcher to build hundreds or even thousands of models that appertain to individual consumer levels. Also, this granularity can lead to coefficients with an unreasonable sign and/or magnitude (Montgomery and Rossi, 1999). Also, many researchers are comfortable with calculating the consumer behaviour to estimate the sales of a product, and hence, the aggregated coefficients of models answer their needs (Oliveira-Castro *et al.*, 2006).

A middle ground may be the consideration of the data structure itself. Buyers form a part of a household, and therefore, there is a hierarchical structure to the data, where purchases are made within the household. It can be seen then, that there is a form of hierarchy in the data since multiple purchases are made by households for every category, with the biscuit category demonstrating the highest average purchases per category for the 52 week period. Therefore, it can be envisaged the structure of the data is a hierarchical one where purchases are made within household, as displayed in Fig 39.

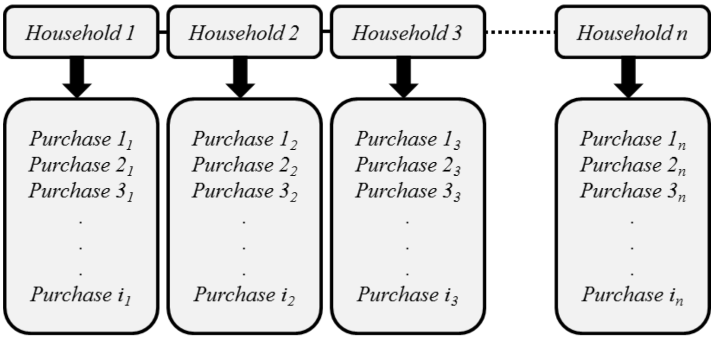

Figure 39: Hierarchical structure of the data

Behaviour of the households becomes conditioned on historic purchase experiences, which is a fundamental input of the Behavioural Perspective Model (e.g. Foxall (2013). The implication places the assumption of independence between individual purchases of the database into question since one can hypothesize that purchases within household may rely on historic conditioning of the category. Therefore, while the assumption of independence of purchases *across* household is realistic, the assumption of independence of purchases *within* household may not be. Not accounting for the hierarchical structure may result in underestimated regression standard errors. This could result in the erroneously determining a statistically significant causality between the independent and dependent variable (Browne and Rasbash, 2004).

The current model is set up assuming that each purchase record in the data file is independent of each other, though in order to test whether a hierarchical structure is more representative of the data then the models will also include a hierarchical element where the hierarchy will be the household. This can be easily identified in the data through the *panel id* variable, since it is unique to a household.

3.8.1 Fixed and Random Effects

A hierarchical model may consist of fixed effects, random effects or both (which is known as a mixed effects model). A fixed effect allows inferences to be made about variables and values specific to the sample within the study, whereas a random effect is used if inferences are to be generalised to a wider group (Field *et al.*, 2012). Within this study, the sample of 1,689 households is a representative sample of the GB population hence, within the hierarchically structured models, a random effect is assigned to these. The simplest form of a random term is that of a random intercept, assuming the intercepts vary across the contextual group, in this case, the household. (Field *et al.,* 2012).

Consider the focal and non-focal variables of this study. While the results of the focal variables may result in working hypotheses for other product categories, the specific results of these focal variables are relevant to the specific category and are not intended to represent a generalization to other product categories available within the GB market place.

Furthermore, the non-focal variables are specific to the categories they represent and cannot be generalised to all GB FMCG categories. Hence the focal and non-focal variables will be represented by a fixed effects parameter.

Chapter 4: Research Questions and Univariate Analysis

4.1 Research Question Development

This next section takes advantage of the literature review, category analysis and initial analysis sections and builds on these to construct a number of research questions to be explored in this study.
From the literature, the concept of analysing each category as a separate entity versus analysis of the categories in one combined model has been discussed. In order to formally test and compare these different models, the research questions will be first considered for separate categories. Following this, research questions relating to the combined category analysis will be discussed and presented. This is done in this way as it presents a logical way of building the analysis.

The literature review demonstrates the benefits of a behavioural consumer analysis approach to consumer understanding, especially when considering actual consumer behaviour versus planned consumer behaviour. Within the behavioural analysis literature, the Behavioural Perspective Model is the most developed in terms of understanding radical behaviourism (Wells, 2014) hence this model is employed as a theoretical framework for the study. The flexibility of the model has allowed numerous studies in multiple categories and geographies (e.g. Foxall, 2016a, b, 2017; Foxall and James, 2003; Foxall and Schrezenmaier, 2003; Foxall *et al.*, 2004; Foxall *et al.*, 2006; Oliveira-Castro *et al.*, 2005; Romero *et al.*, 2006).

The literature suggests the price of the product is an important aspect of the behavioural economic aspects of consumer understanding. The negative relationship between price and purchase behaviour is prevalent in behavioural economics and within the BPM framework (e.g. Oliviero-Castro *et al.*, 2006; Broadbent, 1980; Gabor, 1988; Nagle, 1987; Roberts, 1980; Telser, 1962; Chang, 2007; Foxall *et al.*, 2013). While this is not new research, it is important for any new consumer research to assess whether price remains an important variable in the consumer behaviour model. Also, to omit the price variable will cause the model to attempt to compensate for the effect through other variables within the model, hence distorting the true effect of these variables included within the model. Also, assessing how the price elasticity of demand may, or may not vary under a more complex model will be an interesting addition to the subject. Hence the first research question addresses this area and for each category model, the following research question is considered. The research question is subdivided into each of the four categories.

RQ1: Does the average price of the products within the category influences consumer economic behaviour?

Past studies have shown the psychological variables of the BPM, in terms of the Informational and Utilitarian reinforcement variables are influencing consumer behaviour. This is maintained within this study through the following RQ.

RQ2: Are the BPM psychological variables accounting for consumer behaviour for each category. the nature of the supermarket own brand impacting consumer behaviour of the category through differing behaviour at a consumer psychological level, either at a utilitarian and/or informational reinforcement level?

The literature suggests brands which are considered as a higher equity are considered to have a higher informational and utilitarian reinforcement associated with them and the nature by which the BPM had allocated informational and utilitarian reinforcement scores underline this principle. However, the literature also suggests the prevalence of supermarket own brands may have a different influence on how consumers view the brand. The results of the category analysis show the differing nature of the supermarket own brand in terms of the informational and utilitarian reinforcement responses. This could mean a different strategy is required when marketing and retailing these types of brands.

This research aims to build on previous studies by exploring the nature of the psychological impact of products being formally branded as supermarket own brands and any impact the utilitarian and/or informational reinforcement may have on consumer purchase patterns. The BPM’s flexibility lends a suitable framework for exploring this concept and hence the following category specific research questions are constructed.

RQ3: Is the nature of the supermarket own brand impacting consumer behaviour of the category through differing behaviour at a consumer psychological level, either at a utilitarian and/or informational reinforcement level?

In a similar fashion to RQ3, the seasonal pattern of the Christmas week has a negative effect on total category volume, as seen in the category analysis. However, it is not clear whether this difference is due to changes in consumer psychology attitude within the seasonal period or a more general decrease in category purchase through less consumption and less shopping days during the period. Hence this research aims to test this by seeking to understand whether consumer psychology attitudes to informational and utilitarian reinforcement change within the Christmas period. Hence RQ4 is constructed for each category in turn.

RQ4: Is the seasonal Christmas week impacting consumer behaviour within the category, through different levels of utilitarian and/or informational reinforcement during the Christmas seasonal week?

The next research area focusses on how the structural development of the model itself within the BPM framework. The literature has discussed the potential advantages of a hierarchical structure to the model and this structure is also appealing from a theoretical perspective. The argument is the data follow a hierarchical structure where purchase is located within household, hence questioning the assumption of independence made when modelling the data in a non-hierarchical structure. Therefore, this study will also construct the model within a hierarchical framework using the BPM theoretical framework. This will enable comparisons to be drawn between the model performance and the interpretation of the variables from a hierarchical and non-hierarchical framework. Hence RQ5 is structured as follows:

RQ5: Will the modelling of the category within the BPM structure benefit from a hierarchical model structure? What differences in interpretation would be included versus a non-hierarchical framework?

The nature of the Bayesian estimation employed within this study opens the discussion on what nature the prior distribution should take. The study will incorporate both vague prior distributions and informative prior distributions to ascertain any differences this brings to each of the category models. Hence RQ6 follows.

RQ6: How will Bayesian inference utilizing informative and vague priors impact the predictive nature of the model and the interpretation of the parameters?

The next area of research originates from the analysis of the cross-category consumption observed by households. Households are predominantly purchasing from more than one category during the year. This questions the assumption of independence within household, between category purchases. Hence a better consumer understanding may be obtained through looking at all purchases from all four categories in one combined model.

This combined category may be considered in the form of two structures as discussed in the literature. The pooled structure allocated a parameter value to each variable across category whilst a fixed effect model allocates a parameter value variable within category. The advantages of both methods have been discussed and appropriate research questions are now constructed. The research questions are therefore formulated as set forth:

RQ7: Does a combined category model, incorporating all four categories in one model, utilising a pooled parameter structure help the interpretation of consumer behaviour both from a model diagnostic and interpretation perspective? Or does a combined category model, incorporating all four categories in one model, utilising an offset parameter structure help the interpretation of consumer behaviour both from a model diagnostic and interpretation perspective?

Finally, given the data will be modelled in the first place as four separate categories and thereafter as one combined model, incorporating all categories, it will be interesting to test which structure provides the better model in terms of diagnostics and ease of interpretation. Hence RQ8 is as such.

RQ8: How does the diagnostic measured and parameter estimation differ between treating the data as four separate category models versus one combined cross-category model.

4.2 Univariate analysis

This section seeks to establish whether relationships exist at a univariate level between the dependent variable and each of the independent variables in turn. The dependent variable within each category is the volume sold. The independent variables considered are price per 100g for biscuits and baked beans category, while the fruit juice and yellow fats categories are measured in price per 100ml. The other variables relate to those of the BPM, namely the Informational reinforcement and Utilitarian reinforcement. Finally, each category has been grouped into a product type and a pack size variable as discussed in the methods section. These variables are also considered and whether differences exist amongst them.

4.2.1 Biscuits

The biscuit category is the first category to analyse. Each variable is considered in turn and compared to the dependent variable.

4.2.1.1 Volume

The distribution of the volume variable is shown in Fig 40. The distribution has a positive skew. In order to proceed with Pearson's correlation and ANOVA analysis, these proceadures assume a Gaussian nature to the distribution. Therefore, a natural log transformation is taken. Another reason for this tranformation is the intention of running multiple categories within one model. A natural logarithm will transform the data to change in volume which will then be comparable across categories regardless of units. Fig 40 shows the boxplot of the naturally logged volume variable which is now robustly normally distributed.

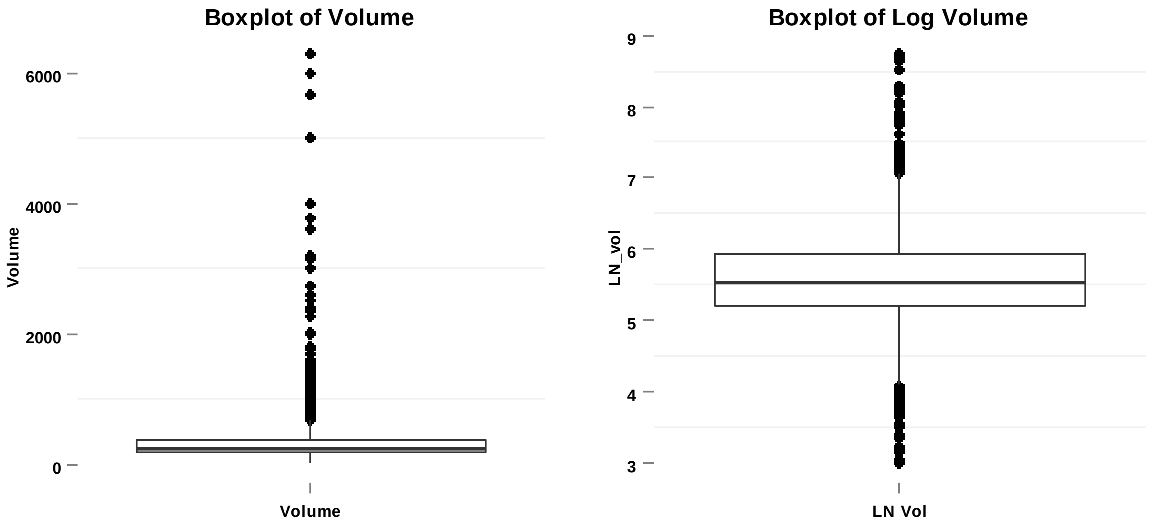

Figure 40: Boxplot of volume and log volume - biscuits

4.2.1.2 Price

The price variable for the biscuit category is defined as the price per 100g and hence can be compared across all formats and pack sizes as discussed in the methods section. The distribution of this variable is shown in Fig 41 and demonstrates a positive skew to the data. A natural log transformation is undertaken on the variable which results in a broadly normal distribution of the data. This trasformation will also allow the price to be comparable to the price variables in other categories if a combined analysis is undertaken since it considers the change in the variable rather than the actual amounts. The boxplot of the naturally logged variable is shown in Fig 41, demonstrating a robustly normal distribution, all be it with a long positive whisker.

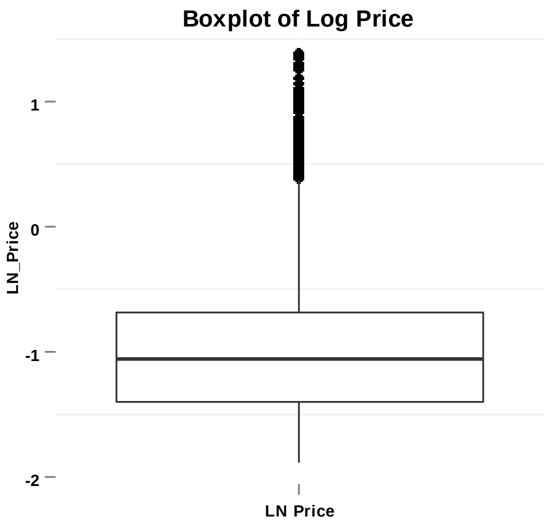

Figure 41: Boxplot of log price - biscuits

To determine whether there is a bivariate relationship between the (log transformed) volume and price variables, a Pearson's correlation analysis is adopted, given the Gaussian nature of the trasnformed variables. It is conceiveable the biscuit volume will be inversly effected by changes in price (e.g. Oliviero-Castro *et al.*, 2006; Broadbent, 1980; Gabor, 1988; Nagle, 1987; Roberts, 1980; Telser, 1962), hence a one-tailed test is undertaken (e.g. Field, 2012). Under the test, the following hypothesis is established.

$H_{biscuits}0_A$: No relationship between log volume and log price

$H_{biscuits}1_A$: There is a negative relationship between log volume and log price

The correlation analsyis gives a correlation coeffcient of -0.601 which is significant at $p<0.0001$ which means that price and volume are statistically significantly negatively correlated, in line with expectations.

4.2.1.3 Informational reinforcement.

Recall from the volume and price variables, a transformation was undertaken for two reasons: First, for the assumption of normality to hold; and second for the variable to be comparable with other categories (in later chapters). With regards to the informational reinforcement variable, Fig 42 shows the variable is robustly normally distributed and hence no transformation to normality is required. Also, since the Informational reinforcement variable is the same scale for all categories, the transformation is not required for compatability between category analysis.

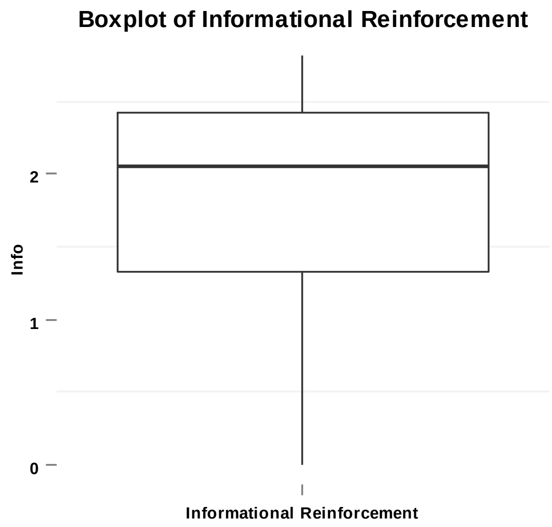

Figure 42: Boxplot of informational reinforcement - biscuits

A univariate test is considered to test the relationship between the two variables and a Pearson's correlation analysis is undertaken. The hypothesis is constructed as before, though this time no assumption is made as to the nature, if any, of the relationship existing between the Informational reinforcement variable and the transformed volume variable. Therefore, a two-tailed test is considered with the following hypothesis being established under the test assumptions.

$H_{biscuits}0_B$: No relationship between log volume and log informational reinforcement
$H_{biscuits}1_B$: There is a relationship between log volume and log informational reinforcement

The returned Pearson's correlation coefficient is 0.166 which is significant at $p<0.0001$ providing sufficient evidence to reject $H0_B$. Therefore, there is a statistically significant correlation between the volume and informational reinforcement variable within this category.

4.2.1.4 Utilitarian Reinforcement

The Utilitarian reinforcement variable is dichotomous in nature, so is not suitable for correlation analysis (Field *et al.*, 2012). The study therefore considers whether the average volume of biscuits purchased is higher within one of the utilitarian reinforcement groups than the other. The mean volume for each group is 5.39 (lower utilitarian reinforcement group) and 5.57 (higher utilitarian reinforcement group). An ANOVA is conducted to test whether this difference is statistically significant. The dependent (volume) variable is logged (given the ANOVA calculations are based on p-values from the student's t distribution, which are robustly normally distributed). As with the informational reinforcement variable, no prior assumption is made as to the nature (if any) of the relationship and hence a two-tailed test is set up. Under the ANOVA, the following hypothesis is established.

> $H_{biscuits}0_C$: Mean level of (naturally logged) volume is the same for each utilitarian reinforcement level
> $H_{biscuits}1_C$: Mean level of (naturally logged) volume is not the same for each utilitarian reinforcement level

The result of the ANOVA yields a high F ratio of 1315, statistically significant at $p<0.0001$, which suggests sufficient evidence to reject $H0_C$ and hence the mean (naturally logged) volume is significantly different between the two utilitarian reinforcement groups, with the higher Utilitarian reinforcement group having the largest volume.

4.2.1.5 Supermarket Own Brand.

From the category analysis section, it has been seen the supermarket own brands may behave in a different manner from the other brands. To test whether the average (transformed) volume is statistically different between the two levels, an ANOVA is employed. The volume variable is logged given the normality assumptions of the test. As with the informational and utilitarian reinforcement variables, no assumption is made as to whether the supermarket own brands have a higher or lower average volume than the non-supermarket own brands and hence a two-tailed test is utilised. Under the test, the following hypothesis is established:

$H_{biscuits}0_D$: Mean level of (naturally logged) volume is the same for the supermarket own brand and branded items.

$H_{biscuits}1_D$: Mean level of (naturally logged) volume is not the same for the supermarket own brand and branded items.

The result of the ANOVA yields a high F-ratio of 118 and a $p<0.0001$ means the null hypothesis $H0_D$ can be rejected. Hence there is a statistically significant difference between the supermarket own brands (mean logged volume of 5.54) and non-supermarket own brands (mean logged volume of 5.48) hence the average volume for supermarket own brands is higher than that of non-supermarket own brands.

4.2.2 Fruit Juice

A similar analysis approach is adopted for the fruit juice category with the same assumptions made as per the biscuit category.

4.2.2.1 Volume

Fig 40 shows the positively skewed volume distribution. A naturally logged transformation is taken and the resulting distribution shown in Fig 43 is robustly normally distributed. This transformation will satisfy any normal assumptions and also allow cross category analysis in subsequent chapters. Fig 43 shows the distribution of the naturally logged volume for this fruit juice category.

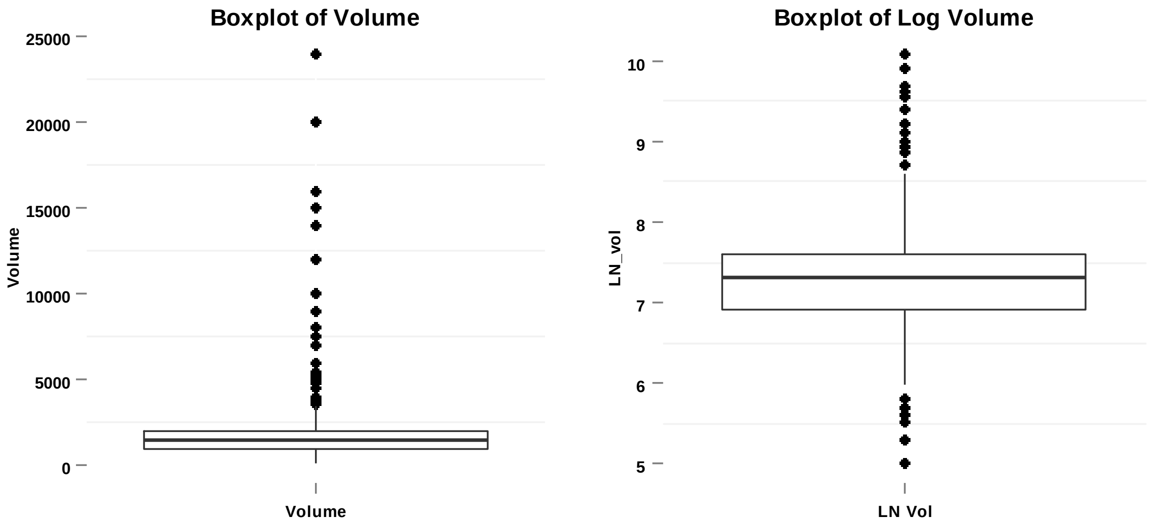

Figure 43: Boxplot of volume and log volume - fruit juice

4.2.2.2 Price

The price of the fruit juice category is calculated as volume per 100ml in order to be comparable across different products and pack sizes. As with the biscuit category, the price variable is transformed to robust normality using the natural log transformation and the boxplot of this variable is shown in Fig 44.

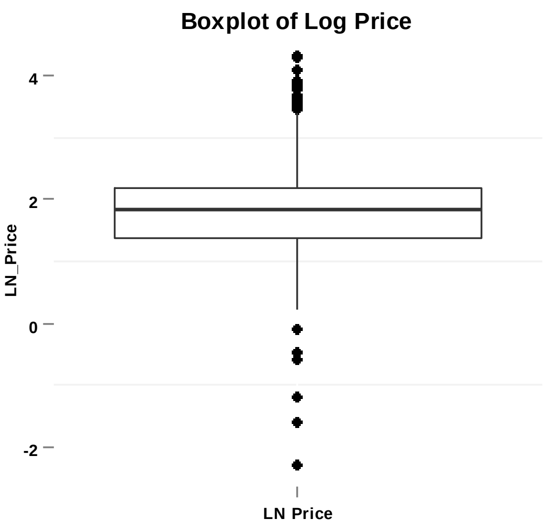

Figure 44: Boxplot of log price - fruit juice

To assess whether there is a relationship between the (naturally logged) volume and price variables, a Pearson's correlation is employed. A one tailed test is selected, based on the assumption that the relationship between price and volume demand will be inversely related. Under the test the following hypothesis is established.

$H_{\text{fruit juice}}0_A$: No relationship between log volume and log price

$H_{\text{fruit juice}}1_A$: There is a negative relationship between log volume and log price

The Pearson's correlation coefficient returned is -0.359 with a significance level of $p<0.0001$, resulting in sufficient evidence to reject $H_{\text{fruit juice}}0_A$. Hence there is a statistically significantly negative relationship between volume and price within the fruit juice category.

4.2.2.3 Informational Reinforcement

As with the biscuit category the informational reinforcement variable is shown to be robustly normally distributed as per Fig 45.

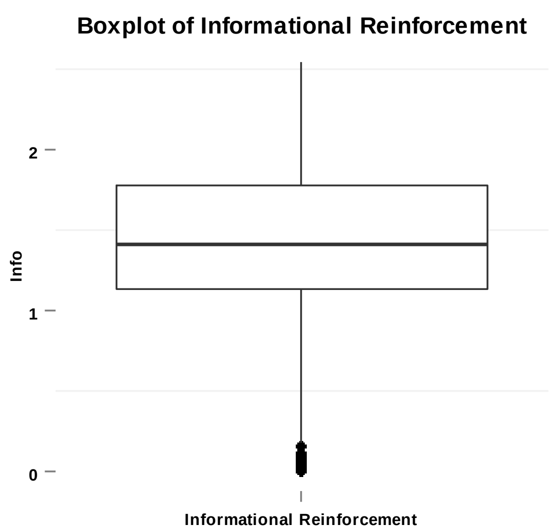

Figure 45: Boxplot of informational reinforcement - fruit juice

A Pearson's correlation analysis is employed on the (log) volume and the informational reinforcement variables to ascertain any relationship between the two. As with the biscuit category and for the same reasons, a two-tail test is employed with no assumptions as to the nature of the relationship (if one exists).

> $H_{\text{fruit juice}}0_B$: No relationship between log volume and Informational reinforcement
> $H_{\text{fruit juice}}1_B$: There is no relationship between log volume and Informational reinforcement

The Pearson's coefficient of the test is 0.014 with a significance level of 0.046, which means there is evidence to reject $H_{\text{fruit juice}}0_B$ at the 5% level. Hence there is a statistically significantly positive relationship between volume and informational reinforcement within this category.

4.2.2.4 Utilitarian Reinforcement

As with the biscuit category, utilitarian reinforcement is a dichotomous variable representing the upper and lower reinforcement levels. The mean volume for each level is 7.42 (lower level) and 7.16 (higher level). In order to formally test this difference, an ANOVA is employed, using the logged volume variable given the underlying Gaussian assumptions of the test. A two-tailed test is employed, making no assumptions as to the nature, if any, of which level may have the highest average volume. Under the test, the following hypothesis is established.

> $H_{\text{fruit juice}}0_C$: Mean level of (naturally logged) volume is the same for each utilitarian level
> $H_{\text{fruit juice}}2_C$: Mean level of (naturally logged) volume is not the same for each utilitarian level

The large F-Ratio (372) results in a highly significant p value ($p<0.0001$) which means significant evidence to reject $H_{\text{fruit juice}}2_C$, hence the average volume within the lower utilitarian level is statistically larger than that of the higher utilitarian reinforcement level.

4.2.2.5 Supermarket Own Brand

Supermarket own brand is a dichotomous variable and hence an ANOVA approach is a preferred option for assessing the mean level of volume between the two groups. The mean of the supermarket own brand is 7.43 and the same statistics for non-supermarket own brand is 7.21. A two-tailed test is established to formally test this difference. No assumption is made of which of the two groups may have a higher average volume and hence a two-tailed test is undertaken. Under the test, the below hypothesis is constructed:

> $H_{biscuits}0_D$: Mean level of (naturally logged) volume is the same for each supermarket own level
> $H_{biscuits}1_D$: Mean level of (naturally logged) volume is not the same for each supermarket own level

The returned F-ratio is 382 ($p<0.0001$) means the $H_{biscuits}0_D$ hypothesis is rejected, meaning the supermarket own brands are accounting for a statistically significantly larger average volume per purchase than the non-supermarket brands.

4.2.3 Yellow Fat

4.2.3.1 Volume

The boxplot of the volume of Yellow Fats category is shown in Fig 46. Its positive skew means a log transformation is taken and the resulting boxplot is shown in Fig 46. From Fig 46, it can be seen the variable remains to be positively skewed since the median and is almost the same value as the lower quartile value, hence the Gaussian assumptions may not be fulfilled in this case. It is still useful to take this transformation, however, in order to meet to the second reason for transformation, i.e. being able to compare the variable across category in future analysis.

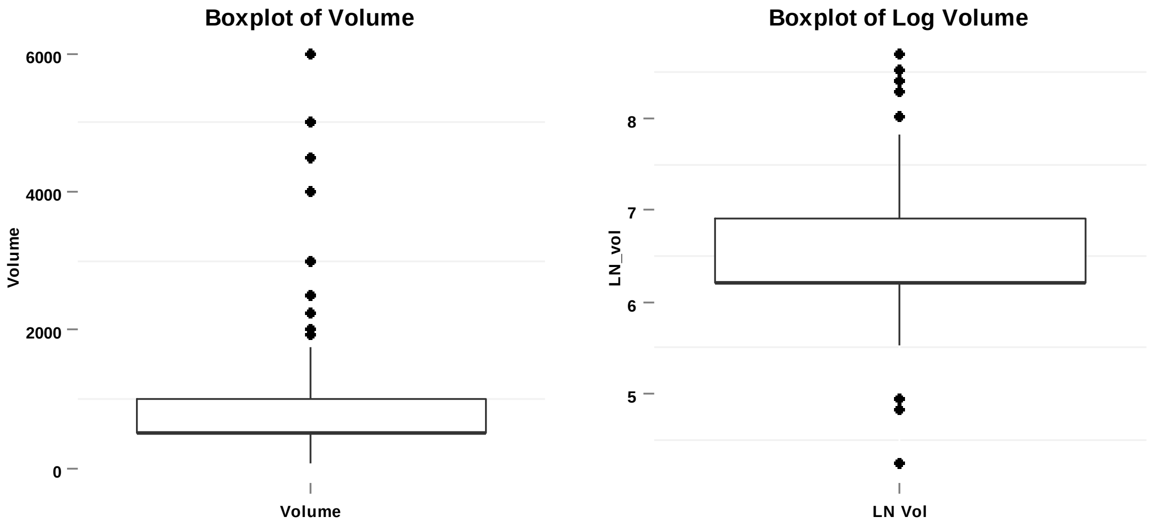

Figure 46: Boxplot of volume and log volume - yellow fats

4.2.3.2 Price

Price is calculated as the price per 100ml in order to be comparable across different fat types and pack sizes. This is then transformed to normality using a natural log transformation and Fig 47 shows a boxplot of the transformed data which can be seen to be of a Gaussian nature. This transformation is made as it can be directly comparable to other categories as the log transformation indicates the change in price.

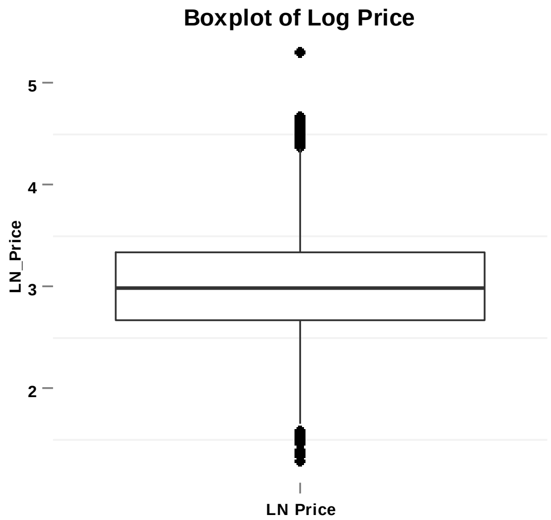

Figure 47: boxplot of log price - yellow fats

To test whether a negative relationship between the volume and price exists, a one-tailed Spearman's correlation analysis is undertaken (one tailed since the assumptions the price demand will be inversely related to price, as per the previous two categories). Spearman's correlation test replaces the Pearson's correlation test, given the non-Gaussian nature of the Log Volume variable. Under the test the following hypothesis is established.

$H_{yellow\ fat}0_A$: No relationship between log volume and log price

$H_{yellow\ fat}1_A$: There is a negative relationship between log volume and log price

The returned Spearman's correlation coefficient is -0.493 ($p<0.0001$) which means evidence that the null hypothesis of $H_{yellow\ fat}0_A$ can be rejected. Therefore, there is a statistically significantly negative relationship between price and volume within the yellow fats category. This is expected and in line with other categories.

4.2.3.3 Informational Reinforcement

The informational reinforcement for yellow fats is shown in the boxplot in Fig 48, demonstrating a robustly normal distribution.

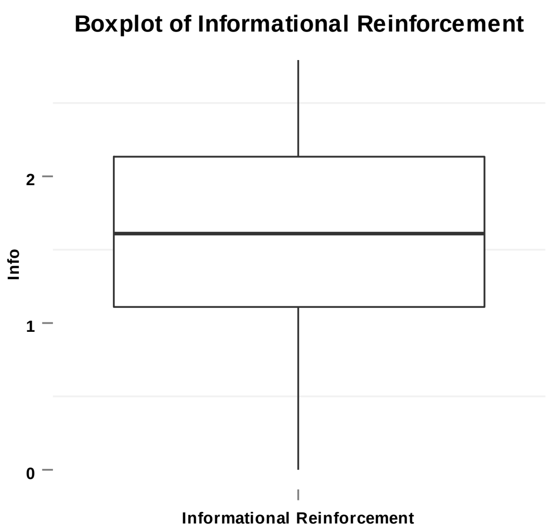

Figure 48: Boxplot of informational reinforcement - yellow fats

For the same reasons as with the price correlation analysis, Spearman's correlation is used instead of a Pearson's. A two-tailed correlation is established, making no assumption of the relationship between the two variables, as per the other categories. The underlying test's hypothesis is thus:

$H_{\text{yellow fat}}0_B$: No relationship between log volume and Informational reinforcement
$H_{\text{yellow fat}}1_B$: There is no relationship between log volume and Informational reinforcement

The Spearman's correlation coefficient of 0.052 ($p<0.0001$) gives sufficient evidence to reject the null hypothesis $H_{\text{yellow fat}}0_B$ and hence demonstrates a statistically significantly positive relationship between volume and informational reinforcement within this category.

4.2.3.4 Utilitarian Reinforcement

The binary utilitarian reinforcement variable is dichotomous and the mean volume for the lower level group is 6.31 and the higher level group is 6.07. Given the non-Gaussian nature of the data which undermines the assumptions of an ANOVA, a Kruskall-Wallis non-parametric test is used to assess whether the mean volume is the same for the two utilitarian reinforcement groups. As per previous categories, no assumption is made as to the nature (if any) of the relationship and hence a two tailed test is employed and the hypothesis as follows.

$H_{\text{yellow fat}}0_C$: Mean level of (naturally logged) volume is the same for each utilitarian level
$H_{\text{yellow fat}}2_C$: Mean level of (naturally logged) volume is not the same for each utilitarian level

The returned mean ranks for the lower and upper utilitarian reinforcement groups are 16060 and 12325 respectively with a large Chi-square value of (asymptotic significance <0.0001), suggesting the lower utilitarian reinforcement group has a statistically significantly higher mean rank than the higher group.

4.2.3.5 Supermarket Own Brand

The supermarket own brand variable is also dichotomous. The mean level of the supermarket own brand is 6.22 and the non-supermarket own brand 6.29. A Kruskall-Wallis test is employed to formally test the difference in these mean rank levels. A two-tailed test is employed making no assumptions as to which may be the highest.

> $H_{yellow\ fats}0_D$: Mean level of (naturally logged) volume is the same for each supermarket own brand and branded items
> $H_{yellow\ fats}1_D$: Mean level of (naturally logged) volume is not the same for each supermarket own brand and branded items.

The resulting mean ranks for the lower and upper utilitarian reinforcement groups are 15713 and 14586 respectively with a large the resulting Chi-Square value of 119 with an asymptotic significance <0.0001 indicates sufficient evidence to reject $H_{yellow\ fats}0_D$. This means the supermarket own brands have a higher mean rank volume and hence implicitly a higher mean volume per purchase than the non-supermarket own brands.

4.2.4 Baked Beans

4.2.4.1 Volume

As with previous categories, volume is positively skewed and hence a natural log transformation is undertaken and Fig 49 shows the results in robustly normally distributed data. This also allows comparability between categories as discussed earlier.

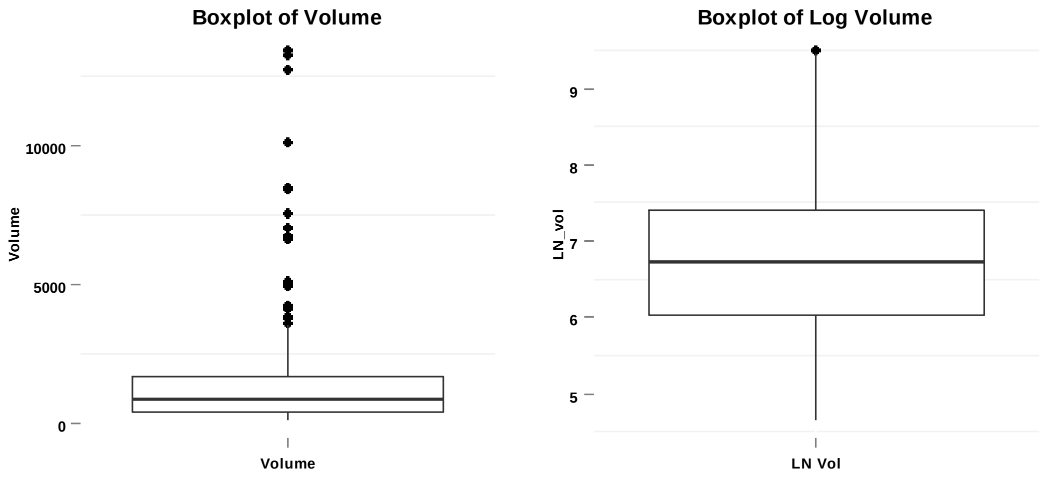

Figure 49: Boxplot of volume and log volume - beans

4.2.4.2 Price

Price is calculated as price per 100g in order to be comparable across products and pack sizes. The price variable is also transformed using the naturally logged function. The resulting distribution is shown in Fig 50 showing the price variable is robustly normally distributed.

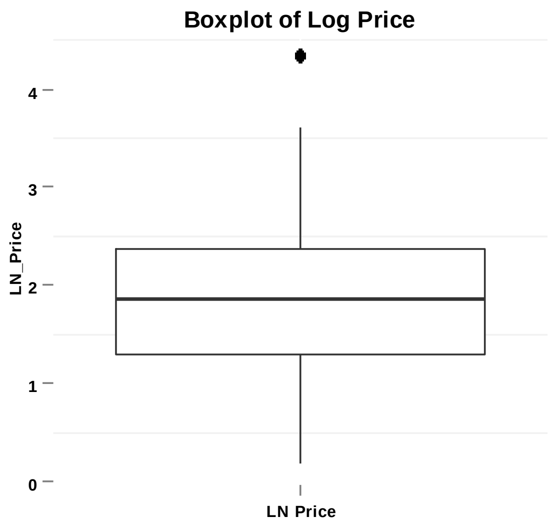

Figure 50: Boxplot of log price

Given the (logged) price and volume variables are robustly normally distributed, a Pearson's correlation analysis is employed. A one tailed test is established given the envisaged inverse relationship between volume demand and price. The hypothesis under the test assumptions is shown below.

$H_{\text{baked beanst}}0_A$: No relationship between log volume and log price

$H_{\text{baked beans}}1_A$: There is a negative relationship between log volume and log price

The Pearson's correlation coefficient is -0.463 ($p<0.0001$) indicating significant evidence to reject the null hypothesis $H_{\text{baked beanst}}0_A$, hence a statistically significantly negative relationship exists between price and volume.

4.2.4.3 Informational Reinforcement

The boxplot of the informational reinforcement variable is shown in the boxplot of Fig 51, demonstrating a robustly normal distribution, though there is no top whisker with the data meaning the top 25% of the data is contained entirely within the top two quartiles,

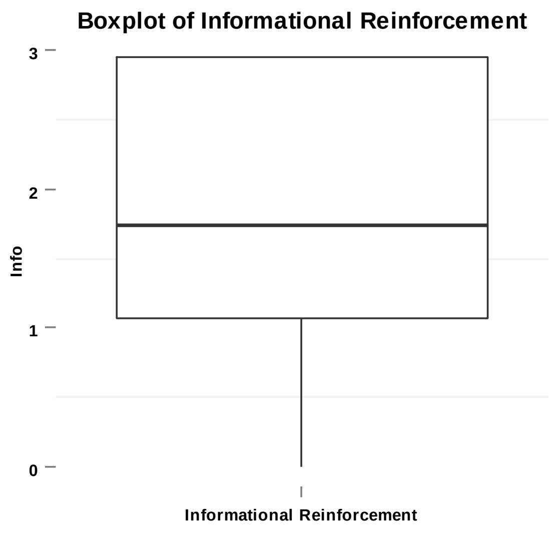

Figure 51: Boxplot of informational reinforcement - beans

The relationship between (log) volume and informational reinforcement is scrutinized using a Pearson's correlation analysis. A two-tailed test is employed with no assumption of the direction of any relationship between the variables. The following hypothesis is therefore established.

> $H_{baked\ beans}0_B$: No relationship between log volume and Informational reinforcement
> $H_{baked\ beans}1_B$: There is no relationship between log volume and Informational reinforcement

The Pearson's correlation coefficient is -0.142 ($p<0.0001$) which gives sufficient reason to reject the null hypothesis $H_{baked\ beans}0_B$. This indicates a statistically significantly negative relationship between volume and informational reinforcement. This is the only category in the analysis which has shown a negative relationship between informational reinforcement and (log) volume.

4.2.4.4 Utilitarian Reinforcement

The utilitarian reinforcement is a binary variable and the mean volume attributed to the lower utilitarian reinforcement group is 6.72 and the higher group 6.35. Given the Gaussian nature of the log volume variable, an ANOVA is set up to formally test whether this mean volume difference is statistically significant between the two utilitarian reinforcement groups. Under the test, the following hypothesis is established.

> $H_{yellow\ fat}0_C$: Mean level of (naturally logged) volume is the same for each utilitarian level
> $H_{yellow\ fat}2_C$: Mean level of (naturally logged) volume is not the same for each utilitarian level

The ANOVA results in a high F-ratio of 665 ($p<0.0001$) where the mean log volume of the lower utilitarian level is 6.722 and the higher level 6.351. Therefore, a statistically significantly higher mean volume for the lower utilitarian reinforcement group.

4.2.4.5 Supermarket Own Brand

The analysis aims to determine whether the mean volume of supermarket own brands is significantly different to non-supermarket own brands. Mean volume for supermarket own brand is 6.71 whereas the mean volume for non-supermarket own brands is 6.50. Supermarket own brand is dichotomous and therefore an ANOVA test is established under the hypothesis.

> $H_{\text{yellow fat}}0_{D}$: Mean level of (naturally logged) volume is the same for each supermarket own brand and branded items.
>
> $H_{\text{yellow fat}}2_{D}$: Mean level of (naturally logged) volume is not the same for each supermarket own brand and branded items.

The resulting ANOVA has an F-ratio of 273 ($p<0.0001$), hence evidence to reject $H_{\text{yellow fat}}0_{D}$ and conclude that supermarket own brands account for a higher mean level of volume than non-supermarket own brands.

4.3 Summary Results

A summary table of the results within this section is displayed in Table 17 to provide a better visualization in the establishing consistencies or differences across category.

		Biscuits	Fruit Juice	Yellow Fats	Baked Beans
Price Correlation	Pearsons Coeff	-.601**	-.359**		-.463**
	Spearmans Coeff			-.493**	
Informational Reinforcement	Pearsons Coeff	.166**	.014*		-.142**
	Spearmans Coeff			.052**	
Utilitarian Reinforcement	Mean log vol (Lower)	5.388	7.420		6.722
	Mean log vol (Higher)	5.574	7.159		6.351
	Mean Rank (Lower)			16060	
	Mean Rank (Higher)			12325	
Supermarket Own Brand	Mean log vol (Super Own)	5.476	7.208		6.498
	Mean log vol (Non Super Own)	5.542	7.429		6.714
	Mean Rank (Lower)			15713	
	Mean Rank (Higher)			14586	

Table 17: Summary of results

From inspection of Table 17, the price correlation coefficient is consistent across the categories in both sign and magnitude. These are also in line with other studies and logical expectations.

In terms of informational reinforcement, there is a significantly positive relationship between (log) volume and informational reinforcement for three of the four categories, with the beans being the exception.

The utilitarian reinforcement variable showed a statistically significant difference between the lower and upper groups in each case. However, the nature of the difference varies between categories. For the biscuit category, the higher utilitarian reinforcement has a higher mean volume than the lower, however for all three other categories it is the other way around. This means that per purchase, brands with lower utilitarian reinforcement have a higher mean volume which may be reflective of a lower price point per purchase enabling larger bulk purchases. This was also evident in the category analysis where the Christmas volume is dominated by the higher utilitarian category.

For supermarket own brands, there is a significant difference between supermarket and non-supermarket own brands. Branded items have a higher mean volume per purchase than supermarket own brands and this time the trend is consistent across all four categories. This lends itself to a similar hypothesis of lower priced brands resulting in larger purchases and a statistically negative elasticity measure would also imply this may be a logical hypothesis.

This analysis indicates the categories are operating within a structure of both behavioural economics and also the Behavioural Perspective Model. The behavioural economics theory can be seen to be enacting on a statistically significant negative relationship between the (log) volume and the (log) price of the product and this can be seen across all categories. Furthermore, there is seen to be a statistically significant relationship between the nature of the brand with differences in behaviour being apparent between supermarket own and non-supermarket own brands.

From a Behavioural Perspective Approach, the informational and utilitarian reinforcement variables are statistically significant in every case which indicates the BPM is influencing purchase decision which has already been discussed by, e.g. Foxall *et al.*, (2011).

4.4 Limitations

Thus far, each variable has been treated as an independent element and assessed in its own right, ignoring any dependencies between the variables. Conceivably there may well be dependencies between the variables and how they interact with each other.
For example, the correlation coefficients between price and informational reinforcement are displayed in Table 18, all of which are statistically significant at ($p<0.0001$).

	Info Reinf Biscuits	Info Reinf Fruit Juice	Info Reinf Yellow Fats	Info Reinf Baked Beans
Price Biscuits	-.122**			
Price Fruit Juice		.339**		
Price Yellow Fats			.367**	
Price Baked Beans				.671**

Table 18: Correlation analysis

There are also relationships between the two categorical variables of utilitarian reinforcement and supermarket-own brand as displayed in Table 18.

In order to account for these potential dependencies, a model structure is required to simultaneously asses these relationships rather than a number of independent statistical tests. The next chapter is the methods chapter and proceeds to discuss how these models may be constructed and evaluated. The chapter reflects on the initial analysis and incorporates information into the modelling structure through employing Bayesian inference techniques.

Chapter 5: Methods

5.1 Introduction

The chapter will begin with a discussion on the philosophical nature as to which this study will be undertaken.
The chapter will continue by explaining the model build in terms of how the variables are constructed and how the output should be interpreted. Initially this will be done on a separate category basis where each category is treated as a completely independent entity. Within each category, three models will be discussed, relating to non-hierarchical model, a hierarchical model with vague prior distributions and a hierarchical model with informative prior distributions with potential advantages and disadvantages of each discussed.

The text will then discuss the potential benefits, statistical and theoretical, of moving to a combined category model whereby all categories are represented in one model. The model is constructed on a non-hierarchical and hierarchical basis.

Finally, the text explains how the Bayesian estimation is set up and run. Also, the model diagnostics are discussed and how they should be interpreted to help formulate model evaluation.

5.2 Philosophy of Science

Understanding the philosophical approach to study is an important aspect of management research since it dictates how the process of data collection and analysis is interpreted (Saunders *et al.*, 2009). Research strategy has a “significant impact on not only what we do but how we understand what we are investigating” (Johnson and Duberley, 2006, p. 108).

“Facts do not exist independently of the medium through which they are interpreted” (May, 2001, p. 28). The way in which reality is viewed is known as its ontology and the knowledge or evidence relating to this reality is known as its epistemology (Mason, 2002; Silverman, 2010). It is the rules by which phenomena is known (Mason, 2002). The assumptions which are made during the collection of the data will inform the methods used to analyse the data and hence inform the interpretation employed. Hence an early clarity on the perspective of the philosophy of research is fundamental.

Marketing management, especially in the Western world is predominantly driven by positivism and is largely deductive in nature (Hirschman, 1986; Badot *et al.*, 2009; Johnson and Duberley, 2011). Also, the marketing research is seen more “businesslike”, which is closer associated with a positivism philosophy (Thomas, 2004, p. 49). However, some scholars disagree with this approach claiming businesses are social entities and a positivistic philosophy is too simplistic (Saunders *et al.*, 2009).

Indeed, in recent years there has been a broader range of philosophical approaches used within the discipline with interpretivism becoming a strong epistemology (Brown, 1993; Marsden and Littler, 1996). Lutz (1989, p. 1) claims consumer research is moving to a constructionist viewpoint “experiencing what Kuhn (1970) identified as a paradigm shift”. Hirschman (1986, p. 238) claims the evolution of marketing is to a more socially constructed way because “knowledge is constructed, not discovered”, while Hanson and Grimmer (2007) suggest this is due to increasing use of qualitative interviews and focus groups within the discipline.
Having more than one perspective within a science should be embraced as the way data is captured and interpreted will be different for each perspective (May, 2001).

Given the importance of understanding the philosophical approach in informing methodology, the areas of positivism and interpretivism are briefly discussed and justifications made as to the choice of philosophical approach to this study.

5.2.1 Positivism

Philosophers of the natural sciences and social sciences note the positivism epistemological and ontological assumption is the “social world exists externally, and that its properties should be measured through objective methods” (Saunders *et al.*, 2009; Easterby-Smith et al, 2011, p. 57). It assumes the observable reality exists and that the social world can be explained by a set of laws in much the same way as the natural world (Saunders *et al.*, 2009).

Auguste Comte first coined the term Positivism in 1853 claiming that human behaviour could be altered under certain conditions and this change in behaviour could be predictable in a value free manner, similar to the natural sciences (Johnson and Duberley, 2011). Comte claimed the predictive laws in the social science allowed humans to alter conditions to gain different results (i.e. cause and effect). This is still very much the positivist management aspect of the science today (Johnson and Duberley, 2011).

Hypotheses relating to people's behaviour can be developed and tested by observing them in an experiment type situation by the gathering of "facts not impressions" (Saunders *et al.*, 2009, p. 114). It assumes that people's reaction to phenomena or situations can therefore be measured, predicted and generalised in much the same way as molecules can be predicted to react in a certain way to the application of heat in a certain situation (May, 2001). Research can be carried out by placing people into a "quasi experimental" environment where elements not wished to be understood can be controlled to isolate and measure the effects on specific variables in question. (May, 2001, p. 10). The future can be based on an inductive argument, with predictions based on the past (Hirschman and Holbrook, 1992).

Positivism can be traced to Plato's absolute truth through objectivity (Johnson and Duberley, 2011). It is a largely post Enlightenment philosophy (Johnson and Duberley, 2011) though the roots are traced to David Hume whose view was against all ideas not based on sensory experiences.

"does it contain any abstract reasoning...does it contain any experimental reasoning...? No. Commit it then to the flames; for it can contain nothing but sophistry and illusion"

(Hume,1748-1975: sec. vii, pt iii cited in Johnson and Duberley, 2011, p. 18).

In the 1920-1930s, a group of philosophers formed the Vienna Circle where they developed logical positivism (Hunt, 1991). This was influenced by Hume, also Wittgenstein's Tracxtus Logico-Philosophicus and Russell's Principia Mathematica (Hunt, 1991). The group were academics mainly from the natural sciences and were "philosophy orientated scientists more than scientifically orientated philosophers" (Hunt, 1991, p. 268) which may account for why Bryman (2008, p. 13) says that "positivism is the application of the methods of the natural sciences to the study of social reality and beyond". Following the rise of Nazism, the group dispersed to the UK and US.

5.2.2 Interpretivism

In the 20th century, Positivism came under attack. In 1959, Karl Popper claimed the death of positivism (Johnson and Duberley, 2011) in his work "The logic of scientific discovery". However, most of the attack has sought to highlight the differences between natural and social sciences, emphasising the role of the human in explaining human behaviour (Johnson and Duberley, 2011).

> *"Human thought is consummately social: social in its origins, social in its functions, social in its forms, social in its application"*
>
> *(Geertz, 1973, p. 360).*

Laing's (1967, p. 53) argument builds on this claiming there should be a distinction between the natural sciences and social sciences in how it is researched: "persons experience the world, whereas things behave in the world".

Different from the positivist, an interpretivist claims people draw on their experiences, discourse and interactions with the environment and from this form their own view of the world (Easterby Smith *et al.*, 2008). Hence, whereas the reality viewed by a positivist is single and composed of discrete elements, to an interpretivist, people construct multiple realities (Hirschman, 1986; Saunders *et al.*, 2009).
Hirschman (1986) sets out the differences between the two philosophies shown in Table 19.

Humanistic Metaphysic	Positivistic Metaphysic
Human beings construct multiple realities	Single reality composed of discrete elements
Researcher and phenomenon mutually interactive	researcher and phenomenon independent
Researcher inquiry directed toward the development of idiographic knowledge	statements of truth that are generizeable across time and context
Phenomenal aspects cannot be segregated into causes and effects	can be segregated into cause and effect
Inquiry inherently value laden	possible and desirable to discover value free objective knowledge.

Table 19: Hirschman's (1986) differences of philosophies

In terms of the experimental nature of the research, whilst positivism suggests the researcher is an observer; constructionist view is the "need to be an emphatic participant observer but also an emphatic participant translator” (Hirschman 1986, p. 240). The human mind has logic of its own and the role of social science is to try and understand it. Understanding the cultures of those interpreting the research is essential to understanding the actors (Laing, 1967). It’s not about the manipulation of several variables but about the “in dwelling” of the researcher. “The researcher and phenomenon... are mutually interactive” (Hirschman, 1986, p. 238). Johnson and Duberley (2011) say an inductive approach to research helps understand these cultures and norms.

In consumer behaviour, the differences in the two principles is that the positivist view sees the marketer as “active” and the consumer as “reactive”, though from a constructionist point of view, both are viewed as “active meaning-makers” (Marsden and Little, 1996, p. 648). The positivist approach places the consumer in a “controlled environment” whereas the constructionist places the consumer in a natural environment.

5.2.3 Adopted Philosophical stance

This study will utilise data collected from household panel data for four FMCG categories. The data has been collected without explicit input from any researcher (the consumer is self-scanning the items purchased and their identity is kept anonymous). Easterby-Smith et al (2011) claims this is the ontology associated with positivism.

The methods employed will be quantitative methods which assumes that behaviour can be measured, modelled and a cause and effect relationship to be derived. It assumes behaviour can be predicted from the understanding of economic, psychological and marketing inputs, in a law-like nature. This constitutes a positivism epistemology (Saunders *et al.*, 2009). Households are selected to form a sample which enables the representation of the wider Great Britain household population. Also, the results gained from the sample of households will be assumed to be generalizable to the wider GB population. This is a philosophical understanding of positivism (May, 2001).

Finally, the study will build on the pre-existing knowledge gained from previous BPM studies and hence contribution will likely be through an incremental nature, which according to Kuhn (1970) is a positivistic way of contributing to knowledge.

Therefore, positivism is an appropriate research philosophy for this study.

5.3 Specifying the Model

The next section will discuss how the model will be specified in light of the literature, initial analysis and philosophical stance.

5.3.1 Separate Model Specification

The data are made up of four categories namely biscuits, fruit juice, yellow fats and beans. Each category model will be modelled in turn with a similar functional form. The variables within the model are divided into two groups. The first is referred to as the *focal* parameters which represent those of most interest to this study, namely the economic, psychology (BPM) and seasonal variables. The *non-focal* variables refer to the flavours and pack sizes which are less of a focus of this study.

5.3.2 Focal Variables

The economic variable will be the average price of the product (price). The price variable is logged which means the value of the coefficient β_1 can be interpreted as the price elasticity of demand. The consumer psychology variables are those of the BPM (see e.g. Foxall, 2013) and refer to the Informational and Utilitarian reinforcement variables. The Informational reinforcement variable is continuous and the Utilitarian reinforcement is dichotomous. In order to assess whether the Informational reinforcement behaves differently for products in different Utilitarian reinforcement groups, an interaction term is constructed and is modelled as an offset variable. Hence a variable representing the Informational reinforcement variable is constructed. An offset variable is then constructed which is equal to the Informational reinforcement if the product has a higher Utilitarian reinforcement group and zero otherwise, or mathematically

$$X_{2i} = I_i$$

$$X_{3i} = \begin{cases} 0 & for \quad U_1 \\ I_i & for \quad U_2 \end{cases}$$

The coefficient β_2 for X_2 then represents the Informational reinforcement within the lower Utilitarian reinforcement group while the coefficient β_3 for X_3 represents the difference attributed to Informational reinforcement within the higher Utilitarian reinforcement group versus the lower Utilitarian reinforcement group. Hence the estimate for the value of the coefficient for the Informational reinforcement within the higher Utilitarian group can be constructed by the addition of both coefficients $\beta_2 + \beta_3$.

It is worth noting that an alternative method of creating the variables would be to create one variable representing the Informational reinforcement within the higher Utilitarian reinforcement group and a second representing the Informational reinforcement within the higher Utilitarian reinforcement group, or mathematically

$$\begin{aligned} X_{2i} &= I_i \quad for \quad U_1 \\ X_{3i} &= I_i \quad for \quad U_2 \end{aligned}$$

The advantage of this is the coefficients of both variables relate directly to the value of the Informational reinforcement variable in the lower and higher Utilitarian reinforcement groups respectively. However, the advantage of the offset method is that the difference between the two coefficients can be evaluated statistically to determine if there is a difference between the two coefficients and hence whether the slope of the Informational reinforcement is performing differently within the lower and higher Utilitarian reinforcement groups.

From the initial analysis, it is deemed that the supermarket own brands may be behaving in a different way from other brands. In order to test this, a dichotomous variable is created which is 1 when the product is a supermarket own brand and 0 otherwise. However, this study seeks to understand how the variable is performing within the theoretical framework of the BPM and specifically whether the Informational and Utilitarian reinforcement variables are behaving differently for supermarket own brands versus other brands. Therefore, a similar approach is undertaken as was discussed with the Informational and Utilitarian reinforcement offsets above. The complication here is the addition of the third variable, namely the supermarket own brand binary variable. Hence two variables are constructed: the first with a resulting estimate β_4 relates to the Informational reinforcement of products which are supermarket own brands. The second variable with an estimated coefficient of β_5 is constructed relating to the Informational reinforcement of supermarket own brands but only for those in the higher Utilitarian group. Mathematically

$$X_{4i} = I_i \quad for \quad S = 1$$
$$X_{5i} = \begin{cases} 0 & for \quad S = 1 \cap U_1 \\ I_i & for \quad S = 1 \cap U_2 \end{cases}$$

This is an offset model as described previously whereby β_4 is the Informational reinforcement for the supermarket own brands within the lower Utilitarian reinforcement group. The β_5 is the offset Informational reinforcement for the supermarket own brands in the higher Utilitarian reinforcement group (to the lower Utilitarian reinforcement group). Hence the estimate for the Informational reinforcement for supermarket own brands within the higher Utilitarian reinforcement group can be derived by the addition of both coefficients $\beta_4 + \beta_5$. The model will be able to determine statistically if the Informational reinforcement for supermarket own brands is statistically different for the Informational reinforcement for non-supermarket own brands. Also, whether the Informational reinforcement for supermarket own brands is statistically different for the higher Utilitarian reinforcement group versus the lower Utilitarian reinforcement group.

Another potential area of interest uncovered in the category analysis chapter, is to understand any behavioural differences during the seasonal Christmas trading week. There is a clear drop in volume for all four categories but it is unclear whether the psychological behaviour within the BPM framework also changes during this extraordinary week in terms of category sales. To analyse the effect, a dichotomous variable is created whereby any transactions within this period is allocated the value 1 and 0 otherwise. In the same manner as discussed for the supermarket own brand variables, the Christmas dichotomous variable is also used in the context of the BPM theoretical framework. Therefore, an interaction term is created between the Christmas binary variable and the Informational reinforcement variable. Similarly, a second variable is created as the interaction of the Christmas dummy variable and the Informational reinforcement variable though only when the Utilitarian reinforcement associated with the product is of the higher group. Thus, is mathematically similar to the supermarket own brand variable construction.

$$X_{6i} = I_i \qquad for \quad Ch = 1$$

$$X_{7i} = \begin{cases} 0 & for \quad Ch = 1 \cap U_1 \\ I_i & for \quad Ch = 1 \cap U_2 \end{cases}$$

Therefore, the coefficient β_6 is the estimate for the Christmas and Informational reinforcement interaction within the lower Utilitarian reinforcement group. The coefficient β_7 is the offset for the Informational reinforcement within the Christmas trading week for the higher Utilitarian group and hence the estimate for the Informational reinforcement within the Christmas trading week for the higher Utilitarian reinforcement group can be derived by adding together $\beta_6 + \beta_7$. As with the previous variables, the offset allows the statistical comparison of the Informational reinforcement of the Christmas week effect between the higher and lower Utilitarian reinforcement groups.

5.3.3 Non Focal Variables

The non-focal variables within this study relate to the flavour or type or product and the number in pack. These are included within the model in order to understand whether the volume of purchase is significantly different between the product variants. The inclusion of these variables ensures a cleaner statistical causality for the focal variables of the parameter. These parameters are the *flavour* and *number in pack* variables

5.3.3.1 Flavour Variable

For each category, the flavour or type of the category is available. Fig 52 below is a list per category, which were introduced in the initial analysis section.

Biscuits				
Chocolat e Coated	Plain Sweet	Filled	Non Sweet	Countlines

Fruit Juice										
Other fruit	Breakfast	Grape	Grapefruit	Mixed	Orange	Pineapple	Tomato	Vegetable	Vitamin	Apple

Yellow Fats			
Butter	Margarine	Low Reduced	Blended spreads

Beans				
Beans Plus	Tomato	Healthy	Flavours	Beans Only

Figure 52: Flavours per category

For each variant within category, a binary variable is created where 1 indicates the product is of that particular variant and 0 indicates the product is not of that variant. Hence each variant within category are mutually exclusive and cumulatively exhaustive (MECE). Since these are MECE, for modelling purposes, one category is chosen as the *base* category and each other variant is an offset to this base. The base variant for each category is indicated in the above table by the shaded box. There are no modelling implications as to which variant is chosen as base. This ensures the underlying matrix is full rank and hence invertible. Therefore, the coefficients of the models will be under the assumption of the base category and the coefficients of each variant will be an offset to the base category. Given the dependent variable is log volume per transaction then the offset will correspond to the difference in mean logged volume per transactions versus the base variant.

5.3.3.2 Numbers in Pack

The number in pack variable is treated in the same way. Fig 53 below shows the variants for this variable across the four categories.

Biscuits					
Size 2-5	Size 6-7	Size 8-11	Size 12+	Size packs	Size 1s

Fruit Juice		
size 2-5	Size 6+	Size 1s

Yellow Fats	
Size 2+	Size 1s

Beans	
Size 4-12	Size 1-2

Figure 53: Packs per category

The structure within the model handles these variables in the same manner whereby each variant is coded as a binary variable and a base variable is elected in order to maintain the full rank matrix. Each variant is then an offset to the *base* variant.

All four categories share a common variable here of one *item per pack* hence there could be an argument to have one base category representing single items and offsets for that for each category. However, given the diversity of category and different purchase cycles, each category is kept separate.

5.3.4 Model Functional Form

The model is therefore constructed with the following functional form shown in Equation 4.

$$
\begin{aligned}
LN(Volume_j) = \beta_0 \\
&+ \beta_1 LN(\Pr ice_j) \\
&+ \beta_2 * Informational_j * Utilitarian_j \\
&+ \beta_3 Informational_j * Utilitarian_{jGroup2} \\
&+ \beta_4 Supermarket * Informational_j \\
&+ \beta_5 Supermarket * Informational_j * Utilitarian_{jGroup2} \\
&+ \beta_6 Christmas * Informational_j \\
&+ \beta_7 Christmas * Informational_j * Utilitarian_{jGroup2} \\
&+ \sum_{i=1}^{a-1} \beta_{i+6} Flavour\_type_i \\
&+ \sum_{i=1}^{b-1} \beta_{i+6+(a-1)} Pack\_type_i \\
&+ \varepsilon_j
\end{aligned}
$$

$$
where\ \varepsilon_j \sim N(0, \sigma^2) \qquad j = 1,2,\ldots,n
$$

Equation 4: Non-hierarchical functional form for separate categories

The model is applied to each category separately and the results and diagnostics discussed in the next Chapter.

5.3.5 Defining the Prior Distributions

The nature of the Bayesian model requires the definition of a prior distribution. As discussed in the literature, the prior distribution is independent of the data and subject to the researcher's disposition.

5.3.5.1 Use of a Vague Prior

As discussed, the use of a vague prior has been used extensively to represent knowledge around a parameter. The study will utilise this prior information around each parameter of the model. This will be referred to as a vague model. The vague prior will be defined from the normal distribution (Lunn *et al.*, 2012). Given the vague nature of the prior, the mean will take the value of zero and the precision 0.001. The same prior distribution will apply to each parameter β of the model and will be of the form

$$
\beta_i \sim N(0, 0.001) \qquad \forall\, i
$$

This will mean the likelihood will have a strong influence on the inference of the posterior distributions of the parameters.

5.3.5.2 Definition of an Informative Prior

As well as the model with *vague* priors, another model is constructed with *informed* priors. This means the researcher will have a degree of influence on the parameter inference since the estimates will be a blend of the informative prior distribution as well as the likelihood from the data. Discussed earlier was the notion of a calibrated prior whereby information taken from frequentist analysis is used to produce the prior distribution itself. This method is adopted for the second model and will allow a comparison on how the results may differ from the use of a vague and informed prior distribution. The estimates for the regression based model are calibrated by running a linear model for each estimate. The mean of the prior becomes the mean of the frequentist linear model. Similarly, the precision of the prior distribution is calculated from the inverse of the squared standard error of the frequentist estimate. Rossi and Allenby (1993) perform a similar procedure by estimating the informed prior from the total MLE estimates of all households. One issue discussed by Dunson (2001) is for large data sets the influence of the likelihood relative to the prior becomes very strong, however since the calibrated prior is estimated from the same large data set, the standard errors of the estimate are relatively small (due to large n) and hence this creates a larger precision which goes to balance the influence of the likelihood somewhat. Table 20 shows the mean, standard error and precision of each focal variable within the four categories of the study.

		Biscuits	Fruit Juice	Yellow Fats	Beans
Price	Beta	-0.72	-0.42	-0.47	-0.55
	Std Error	0.00	0.01	0.01	0.01
	Precission	67061	17931	36674	12147
	Precision / Beta	93493	42725	77754	21940
Informational	Beta	0.14	0.02	0.03	-0.12
	Std Error	0.00	0.01	0.00	0.01
	Precission	91750	15130	44412	20994
	Precision / Beta	667891	932710	1703141	181200
Informational Utilitarian Gp2	Beta	0.11	-0.17	-0.12	-0.24
	Std Error	0.00	0.01	0.00	0.01
	Precission	192052	12363	72441	19729
	Precision / Beta	1817837	73607	627632	83277
Supermarket Own x Informational	Beta	0.04	0.10	-0.03	0.11
	Std Error	0.00	0.01	0.01	0.01
	Precission	88794	22481	30222	11551
	Precision / Beta	2141987	232965	918031	109928
Supermarket Own x Informational Ut 2	Beta	0.09	-0.22	-0.14	-0.09
	Std Error	0.01	0.02	0.01	0.02
	Precission	16016	3696	5229	3141
	Precision / Beta	182354	16582	37955	35523
Christmas	Beta	0.05	0.03	-0.08	0.01
	Std Error	0.03	0.04	0.03	0.06
	Precission	1175	577	1189	262
	Precision / Beta	23625	21974	14438	19208
Chrsitmas Utilitarian Gp2	Beta	0.14	-0.17	-0.22	0.02
	Std Error	0.04	0.13	0.07	0.13
	Precission	511	57	215	62
	Precision / Beta	3624	344	971	2869

Table 20: Informative prior distribution statistics

The ratio of the magnitude of the precision to the absolute magnitude of the estimate (Beta) varies considerably across parameter. For example, within the fruit juice category the ratio of the Precision/Beta for the Informational parameter is 932710, implying the precision is very high and will have a large influence on the estimate of the parameter. In contrast, for the same category of fruit juice, the ratio of Precision/Beta for the Christmas Utilitarian Grp 2 parameter is 344, significantly lower and hence less influence of the prior on the posterior estimate. Similar differences can be found by inspection of the other categories, demonstrating the importance of the prior distribution in terms of its influence on the posterior parameter estimate, which is what was discussed earlier in the literature review.

The above Beta and precision estimates are hence translated into the informed prior distributions of the four categories as shown below in Fig 54. For each model the non-focal variables are of less importance and a vague prior will be used in these cases.

Biscuits
Price ~ N(-0.717, 67061)
Informational ~ N(0.137, 91750)
Informational Utilitarian Gp2 ~ N(0.106, 192052)
Supermarket Own x Informational ~ N(0.041, 88794)
Supermarket Own x Informational Ut 2 ~ N(0.088, 16016)
Christmas ~ N(0.05, 1175)
Chrsitmas Utilitarian Gp2 ~ N(0.141, 511)

Fruit Juice
Price ~ N(-0.42, 17931)
Informational ~ N(0.016, 15130)
Informational Utilitarian Gp2 ~ N(-0.168, 12363)
Supermarket Own x Informational ~ N(0.097, 22481)
Supermarket Own x Informational Ut 2 ~ N(-0.223, 3696)
Christmas ~ N(0.026, 577)
Chrsitmas Utilitarian Gp2 ~ N(-0.167, 57)

Yellow Fats
Price ~ N(-0.472, 36674)
Informational ~ N(0.026, 44412)
Informational Utilitarian Gp2 ~ N(-0.115, 72441)
Supermarket Own x Informational ~ N(-0.033, 30222)
Supermarket Own x Informational Ut 2 ~ N(-0.138, 5229)
Christmas ~ N(-0.082, 1189)
Chrsitmas Utilitarian Gp2 ~ N(-0.221, 215)

Beans
Price ~ N(-0.554, 12147)
Informational ~ N(-0.116, 20994)
Informational Utilitarian Gp2 ~ N(-0.237, 19729)
Supermarket Own x Informational ~ N(0.105, 11551)
Supermarket Own x Informational Ut 2 ~ N(-0.088, 3141)
Christmas ~ N(0.014, 262)
Chrsitmas Utilitarian Gp2 ~ N(0.022, 62)

Figure 54: Informative prior distributions

These Prior distributions will need to be included within the model functional form specification and each model will have two versions of the prior distribution. One version will have a vague prior associated with each focal and non-focal parameter and the other will have assigned informative priors for the focal variables (non-focal variables will remain with a vague prior distribution). An example of each of the variants is shown below utilising the biscuit category as the example. Equation 5 is the vague prior model and Equation 6 is the informative model.

$$
\begin{aligned}
LN(Volume_j) = \beta_0 \\
&+ \beta_1 LN(\Pr ice_j) \\
&+ \beta_2 * Informational_j * Utilitarian_j \\
&+ \beta_3 Informational_j * Utilitarian_{jGroup2} \\
&+ \beta_4 Supermarket * Informational_j \\
&+ \beta_5 Supermarket * Informational_j * Utilitarian_{jGroup2} \\
&+ \beta_6 Christmas * Informational_j \\
&+ \beta_7 Christmas * Informational_j * Utilitarian_{jGroup2} \\
&+ \sum_{i=1}^{a-1} \beta_{i+6} Flavour\_type_i \\
&+ \sum_{i=1}^{b-1} \beta_{i+6+(a-1)} Pack\_type_i \\
&+ \varepsilon_j
\end{aligned}
$$

$$\beta_0 = U_0 + \upsilon_0$$
$$\upsilon_0 \sim N(0, \sigma^2)$$
$$where\ \varepsilon_j \sim N(0, \sigma^2) \qquad j = 1,2,\ldots,n$$

$$\beta_1 \sim N(0, 0.001)$$
$$\beta_2 \sim N(0, 0.001)$$
$$\beta_3 \sim N(0, 0.001)$$
$$\beta_4 \sim N(0, 0.001)$$
$$\beta_5 \sim N(0, 0.001)$$
$$\beta_6 \sim N(0, 0.001)$$
$$\beta_7 \sim N(0, 0.001)$$
$$\beta_i \sim N(0, 0.001) \quad \text{for i} = 8,\ldots,8 + (\text{a} - 1) + (\text{b} - 1)$$
$$\upsilon_0[k] \sim N(0, 0.001) \quad \text{for } k = 1,\ 2,\ldots,h \quad \text{where } h = \#\,\text{households}$$

Equation 5: Hierarchical model with vague priors

$$
\begin{aligned}
LN(Volume_j) = \beta_0 \\
&+ \beta_1 LN(\Pr ice_j) \\
&+ \beta_2 * Informational_j * Utilitarian_j \\
&+ \beta_3 Informational_j * Utilitarian_{jGroup2} \\
&+ \beta_4 Supermarket * Informational_j \\
&+ \beta_5 Supermarket * Informational_j * Utilitarian_{jGroup2} \\
&+ \beta_6 Christmas * Informational_j \\
&+ \beta_7 Christmas * Informational_j * Utilitarian_{jGroup2} \\
&+ \sum_{i=1}^{a-1} \beta_{i+6} Flavour\_type_i \\
&+ \sum_{i=1}^{b-1} \beta_{i+6+(a-1)} Pack\_type_i \\
&+ \varepsilon_j
\end{aligned}
$$

$$
\begin{aligned}
&\beta_1 \sim N(-0.717, 67061) \\
&\beta_2 \sim N(0.137, 91750) \\
&\beta_3 \sim N(0.106, 192052) \\
&\beta_4 \sim N(0.041, 88794) \\
&\beta_5 \sim N(0.088, 16016) \\
&\beta_6 \sim N(0.05, 1175) \\
&\beta_7 \sim N(0.141, 511) \\
&\beta_i \sim N(0, 0.001) \quad \text{for i} = 8,\ldots,8+(\text{a}-1)+(\text{b}-1)
\end{aligned}
$$

$$
where\, \varepsilon_j \sim N(0, \sigma^2) \qquad j = 1,2,\ldots,n
$$

Equation 6: Non-Hierarchical model with informative priors (biscuits)

5.3.6 Structure of the Models

In order to understand how the hierarchical structure of the data may affect the model estimation, non-hierarchical and hierarchical models are built using the household as the random error term.

The hierarchical term is built into the existing non-hierarchical models and the model structure is shown below. The prior distribution also needs to account for the new parameter introduced. The model is run in its vague form and can be directly compared to the vague non-hierarchical model.

5.3.7 Prior Distribution of the Model Variance Term

In both the non-hierarchical and hierarchical structure, the variance coefficient requires a prior distribution. The variance is non-negative of course, hence the normal distribution is not suitable since a non-positive value may be sampled from the distribution which is an absurd value for the variance term as it is a squared term. This would cause the MCMC to produce an error. Therefore, a Gamma distribution is better suited for the prior distribution since it will return only positive values (Spiegelhalter *et al.*, 2002). There are no logical informative values for the variance term and hence a vague prior is constructed for both the variance term for the model τ (i.e. the variance across household) and also the hierarchical variance term τ_v (i.e. the variance between household). The prior distributions, therefore are created as below.

$$\tau \sim Gamma(0.0001, 0.0001)$$
$$\tau_v \sim Gamma(0.0001, 0.0001)$$

5.3.8 Combining categories

As previously discussed, the data are organised by household and "panel id" is a unique identifier of each household. Within the time frame of the data (a calendar year) a household may purchase more than one of the categories. Fig 55 shows the distribution of the percentage of households buying either one, two, three or four of the categories in question and also the corresponding percentage of the purchases made.

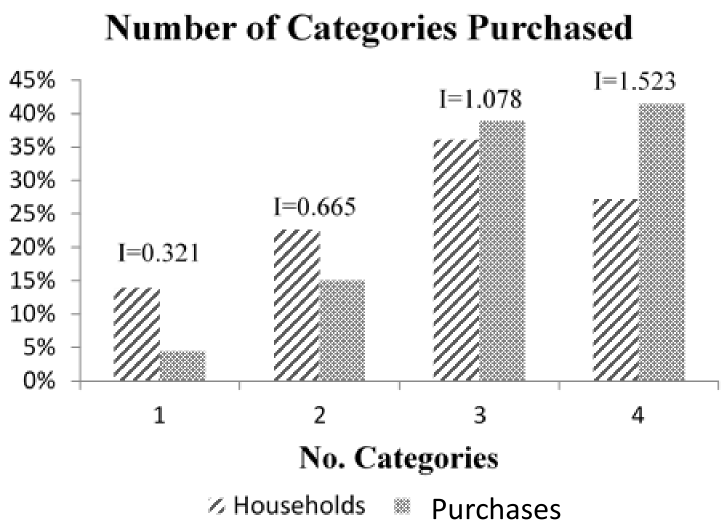

Figure 55: Number of categories purchased per household

This means that 86% of households within the total defined sample purchase more than one category. The purchases made by these 86% of households account for 96% of all items purchased (within the defined sample).
The index represented by the letter "I" is the ratio of the percent of purchases divided by the percent of households, or formally

$$I = \frac{\%\ \text{Purchases}}{\%\ \text{Households}}$$

Where an index of 1 would suggest the proportion of purchases to households is equal. However it can be seen that households who purchase from fewer categories a year tend to be lighter buyers with the index increasing systematically with the more categories that are purchased. This demonstrates a DJ type effect where households with smaller cross-category repertoires also purchase less products within the fewer categories shopped.

5.3.9 Understanding purchase behaviour across the four categories

The number of items purchased in each category can be profiled based on the purchases made as per Fig 56. Biscuits tend to display the largest number of items purchased within a year regardless how many categories are purchased by the household.

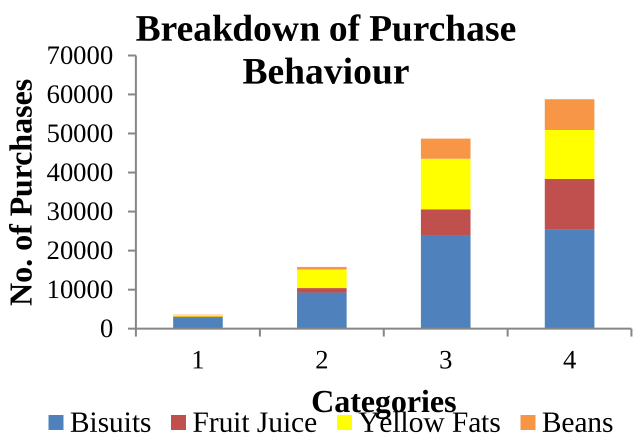

Figure 56: Breakdown of categories purchased per household

The chart above is a summary per number of purchases. The layered pie chart in Fig 57 shows the various combinations of categories purchased within the period split into how many of the categories were purchased in total. The size of each pie is relative to the size of the number of items purchased. The dominance of households who purchase all four categories is clear. Also, the area of the pie charts which contain no blue (i.e. no biscuit purchases) are very small, again highlighting the dominance of the biscuit item purchases.

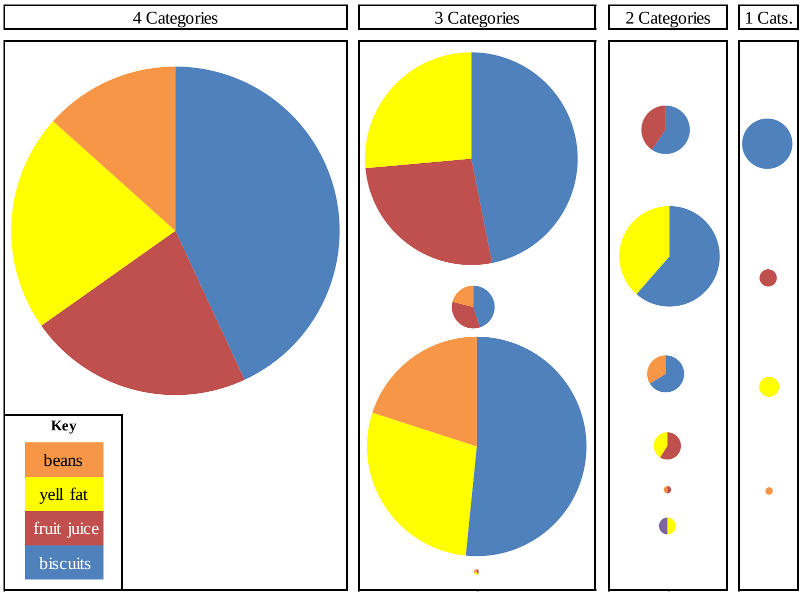

Figure 57: Layered pie chart diagram distribution of categories purchased

Another pictorial view of the purchasing dynamics can be shown in a four way Venn diagram as per Fig 58. The dominance of the biscuit item purchase can be visually identified by the larger numbers appearing throughout the biscuit oblong shape. Yellow fats also show its scale with large values throughout its oblong. The largest value is the value within the centre of the Venn diagram which again shows that the larger purchase bucket is those where all four categories are purchased within the period in question.

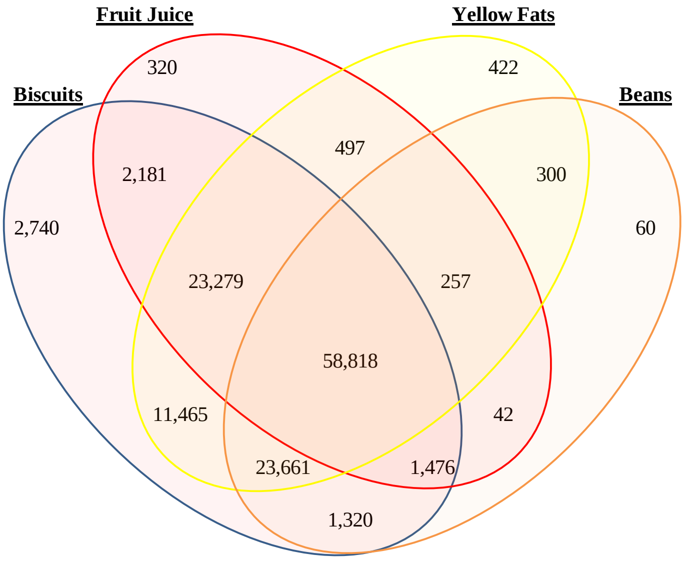

Figure 58: 4-way Venn diagram of category purchase interaction

Another way of viewing the cross category purchase behaviour is to calculate the proportion of households who cross purchase with other categories. Fig 59 shows that given the households who have purchased category *i*, the % of households who have purchased *j, k or l*. It can be seen that within the period in question, of the households who purchased fruit juice, yellow fats and beans, a clear majority of them also purchased from the biscuit category. Yellow fats category is consistently the next largest category given purchases of biscuits, fruit juice or beans. Fruit Juice and Beans are similar with around 50% of households who bought one of the other three categories within the period, also bought fruit juice or beans respectively.

Of the households who bought… ...x% also bought…

	Biscuits	Fruit Juice	Yellow Fats	Beans
Biscuits	100.0%	53.1%	81.0%	50.7%
Fruit Juice	94.5%	100.0%	86.4%	55.1%
Yellow Fats	95.3%	57.1%	100.0%	56.4%
Beans	97.1%	59.3%	91.9%	100.0%

Figure 59: Conditional purchase distribution

5.4 Combined Model

Thus far the four categories have been modelled as completely independent entities. The analysis above however, suggests consumers are likely to purchase across multiple categories during the year and therefore the behaviour of each category may not be independent as the consumers may be showing similar behavioural psychology and economic traits across category. The behaviour within the categories therefore may be influenced by the type of household shopping and their frequency of shopping the category. This is an extension to the argument presented earlier whereby purchases within household may not be independent.

The dependent variables of the BPM are constructed in the same fashion for all categories hence can be interpreted with the same model functional form. The price variables are logged in each case and hence all refer to the change in volume and therefore can also be used within the same model form since the coefficients of each price variable can be compared in this case without further transformation. The dependent volume variable is also logged and hence is also comparable across category. This means the focal variables are all comparable across the model. The flavours and pack size variables are specific to a category. Pack size may sometimes be similar; however, given the differing nature of the categories it does not seem sensible to assume these are comparable. Hence for the non-focal variables, these are kept separate by creating a 0 value for non-relevant categories.

5.4.1. Pooled Structure

A pooled model structure can be used when the variables in questions are relatively homogeneous and the product groups are similar (Joseph, 2010) which is the case in this instance. Within a pooled structure, the dependent variable and the focal independent variables can be stacked into the same variable given their comparable nature (Cameron and Trivedi, 2010). The model structure for these would resemble the following shown in Equation 7.

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \cdots \\ Y_n \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \\ \cdots \\ X_n \end{bmatrix} \beta + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_n \end{bmatrix} = X_i \beta + \varepsilon_i$$

Equation 7: Pooled model structure

However, the reliability of the estimate β will decrease as the homogeneity between the categories decreases (Bass and Wittink, 1975). In this model, each variable will be associated with one parameter estimate which will represent the variable across the four product groups. If in reality the differences between the βs are relatively small between the categories, then Wallace (1972) claims this is a good trade off of maximising the degrees of freedom. However if the differences in the β are significant then the estimates "lack meaning" In that the model may be able to fit a functional form to the data, however the interpretation of the β for management decision making will be a form of average across category which is difficult to use in a practical manner (Bass and Wittink, 1975 p. 414). This can result in a generalised coefficient and limited insights can be drawn across the dataset (Montgomery and Rossi, 1999) which Duncan *et al.,* (1996, p. 819) says is akin to explaining "everything in general and nothing in particular". Another issue with a pooled model is the estimates can be biased though this can be accounted for by introducing a random effects term to account for any inter household dependency (Cameron and Trivedi, 2010).

The non-focal variables have been split into a number of binary variables (see Methods chapter earlier). Since they only apply to a specific category the values outside of the relevant category will be 0.

The pooled functional form of the combined category model is constructed with both a non-hierarchical and a hierarchical structure where household id is used as a random intercept term in the same manner as the model was constructed within the separate model section. Both models are shown below in Equation 8 (non-hierarchical) and Equation 9 (hierarchical).

$$\begin{aligned}
LN(Volume_j) = \beta_0 & \\
& + \beta_1(\text{offset for fj}) \\
& + \beta_2(\text{offset for yf}) \\
& + \beta_3(\text{offset for beans}) \\
& + \beta_4 LN(Price_j) \\
& + \beta_5 * Informational_j * Utilitarian_j \\
& + \beta_6 Informational_j * Utilitarian_{jGroup2} \\
& + \beta_7 Supermarket * Informational_j \\
& + \beta_8 Supermarket * Informational_j * Utilitarian_{jGroup2} \\
& + \beta_9 Christmas * Informational_j \\
& + \beta_{10} Christmas * Informational_j * Utilitarian_{jGroup2} \\
& + \ldots
\end{aligned}$$

$$
\begin{aligned}
LN(Volume_j) = \ldots.& \\
&+ \beta_{11}\text{Chocolate coated (bis)} \\
&+ \beta_{12}\text{Plain Sweet (bis)} \\
&+ \beta_{13}\text{Filled (bis)} \\
&+ \beta_{14}\text{Non Sweet (bis)} \\
&+ \text{Countlines (bis)} \qquad \textit{base} \\
&+ \beta_{15}\text{Size2 - 5 (bis)} \\
&+ \beta_{16}\text{Size6 - 7 (bis)} \\
&+ \beta_{17}\text{Size8 - 11 (bis)} \\
&+ \beta_{18}\text{Size12 + (bis)} \\
&+ \beta_{19}\text{Size packs (bis)} \\
&+ \text{Size1s (bis)} \qquad \textit{base} \\
&+ \beta_{20}\text{Other fruit (fj)} \\
&+ \beta_{21}\text{Breakfast (fj)} \\
&+ \beta_{22}\text{Grape (fj)} \\
&+ \beta_{23}\text{Grapefruit (fj)} \\
&+ \beta_{24}\text{Mixed (fj)} \\
&+ \beta_{25}\text{Orange (fj)} \\
&+ \beta_{26}\text{Pineapple (fj)} \\
&+ \beta_{27}\text{Tomato (fj)} \\
&+ \beta_{28}\text{Vegetable (fj)} \\
&+ \beta_{29}\text{Vitamin (fj)} \\
&+ \text{Apple (fj)} \qquad \textit{base} \\
&+ \beta_{30}\text{Size2 - 5 (fj)} \\
&+ \beta_{31}\text{Size6 + (fj)} \\
&+ \text{Size1s (fj)} \qquad \textit{base} \\
&+ \beta_{32}\text{Butter (yf)} \\
&+ \beta_{33}\text{Margarine (yf)} \\
&+ \beta_{34}\text{Low Reduced (yf)} \\
&+ \text{Blended Spreads (yf)} \qquad \textit{base} \\
&+ \beta_{35}\text{Size2 + (fj)} \\
&+ \text{Size1s (fj)} \qquad \textit{base} \\
&+ \beta_{36}\text{Beans Plus (beans)} \\
&+ \beta_{37}\text{Tomato (beans)} \\
&+ \beta_{38}\text{Healthy (beans)} \\
&+ \beta_{39}\text{Flavours (beans)} \\
&+ \text{Beans Only (beans)} \qquad \textit{base} \\
&+ \beta_{40}\text{Size4 - 12 (beans)} \\
&+ \beta\ \text{Size1 - 2 (beans)} \qquad \textit{base} \\
&+ \varepsilon_j \\
where\ \varepsilon_j \sim N(0, \sigma^2)&
\end{aligned}
$$

Equation 8: Pooled model structure (non-hierarchical)

$$
\begin{aligned}
LN(Volume_j) = \beta_0 \\
&+ \beta_1(\text{offset for fj}) \\
&+ \beta_2(\text{offset for yf}) \\
&+ \beta_3(\text{offset for beans}) \\
&+ \beta_4 LN(\Pr ice_j) \\
&+ \beta_5 * Informational_j * Utilitarian_j \\
&+ \beta_6 Informational_j * Utilitarian_{jGroup2} \\
&+ \beta_7 Supermarket * Informational_j \\
&+ \beta_8 Supermarket * Informational_j * Utilitarian_{jGroup2} \\
&+ \beta_9 Christmas * Informational_j \\
&+ \beta_{10} Christmas * Informational_j * Utilitarian_{jGroup2} \\
&+ \ldots
\end{aligned}
$$

$$
\begin{aligned}
LN(Volume_j) = {} & .... \\
& + \beta_{11}\text{Chocolate coated (bis)} \\
& + \beta_{12}\text{Plain Sweet (bis)} \\
& + \beta_{13}\text{Filled (bis)} \\
& + \beta_{14}\text{Non Sweet (bis)} \\
& + \text{Countlines (bis)} && \textit{base} \\
& + \beta_{15}\text{Size2 - 5 (bis)} \\
& + \beta_{16}\text{Size6 - 7 (bis)} \\
& + \beta_{17}\text{Size8 - 11 (bis)} \\
& + \beta_{18}\text{Size12 + (bis)} \\
& + \beta_{19}\text{Size packs (bis)} \\
& + \text{Size1s (bis)} && \textit{base} \\
& + \beta_{20}\text{Other fruit (fj)} \\
& + \beta_{21}\text{Breakfast (fj)} \\
& + \beta_{22}\text{Grape (fj)} \\
& + \beta_{23}\text{Grapefruit (fj)} \\
& + \beta_{24}\text{Mixed (fj)} \\
& + \beta_{25}\text{Orange (fj)} \\
& + \beta_{26}\text{Pineapple (fj)} \\
& + \beta_{27}\text{Tomato (fj)} \\
& + \beta_{28}\text{Vegetable (fj)} \\
& + \beta_{29}\text{Vitamin (fj)} \\
& + \text{Apple (fj)} && \textit{base} \\
& + \beta_{30}\text{Size2 - 5 (fj)} \\
& + \beta_{31}\text{Size6 + (fj)} \\
& + \text{Size1s (fj)} && \textit{base} \\
& + \beta_{32}\text{Butter (yf)} \\
& + \beta_{33}\text{Margarine (yf)} \\
& + \beta_{34}\text{Low Reduced (yf)} \\
& + \text{Blended Spreads (yf)} && \textit{base} \\
& + \beta_{35}\text{Size2 + (fj)} \\
& + \text{Size1s (fj)} && \textit{base} \\
& + \beta_{36}\text{Beans Plus (beans)} \\
& + \beta_{37}\text{Tomato (beans)} \\
& + \beta_{38}\text{Healthy (beans)} \\
& + \beta_{39}\text{Flavours (beans)} \\
& + \text{Beans Only (beans)} && \textit{base} \\
& + \beta_{40}\text{Size4 - 12 (beans)} \\
& + \beta\ \text{Size1 - 2 (beans)} && \textit{base} \\
& + \varepsilon_j
\end{aligned}
$$

$$\beta_0 = U_0 + v_0$$

$$v_0 \sim N(0, \sigma^2)$$

$$where\ \varepsilon_j \sim N(0, \sigma^2)$$

Equation 9: Pooled model structure (hierarchical)

The prior distributions for the pooled non-hierarchical model are defined as per Equation 10. Recall the combined category utilises vague prior distributions for the reasons discussed earlier.

$$\beta_i \sim N(0, 0.001)$$
$$\text{for } i = 1, 2, \ldots, 40$$
$$\sigma \sim Gamma(0.0001, 0.0001)$$

Equation 10: Pooled prior distributions for the non-hierarchical model

Similarly, the hierarchical model's vague prior distributions are set as per Equation 11 below.

$$\beta_i \sim N(0, 0.001)$$
$$\text{for } i = 1, 2, \ldots, 61$$

$$\upsilon_0[k] \sim N(0, 0.001)$$
$$\text{for } k = 1,\ 2, \ldots, 1689$$

$$\sigma \sim Gamma(0.0001, 0.0001)$$
$$\nu_0 \sim Gamma(0.0001, 0.0001)$$

Equation 11: Pooled prior distributions for the hierarchical model

5.4.2. Fixed Effects Model

A pooled model may be inconsistent if a fixed effects model is more appropriate representation of the data (Cameron and Trivedi, 2010). A fixed effects model allows for the variation of a parameter estimate across groups and hence there will be a separate estimate of β for each group (or in this for each of the four product categories). The β for each category is set up within the model as an offset to the base category, as discussed earlier. This offset structure means that one category will form the base category of an estimate and the other three estimates will be an offset to this. The disadvantage is that the actual estimate of the parameter must be calculated using both coefficients, however the advantage is that inference measures are directly available from the model output to assess whether the differences between the category based estimates are statistically different (i.e. are the offset coefficients statistically different to zero). The advantages of the ability to statistically evaluate the difference of the offset outweighs the disadvantages in the opinion of the author.

5.4.2.1 Price variable Construction for the Fixed Effects Combined Model

Using the offset approach to model construction, the price coefficient estimates are set up as follows (Equation12) where the *biscuits* category is used as the base category.

$$
\begin{aligned}
\text{Price}_{biscuits} &= \beta_{5biscuits} \\
\text{Price}_{fruitjuice} &= \beta_{5biscuits} + \beta_{6\,fruitjuice} \\
\text{Price}_{yellowfats} &= \beta_{5biscuits} + \beta_{7\,yellowfats} \\
\text{Price}_{beans} &= \beta_{5biscuits} + \beta_{8beans}
\end{aligned}
$$

Equation 12: Price variable construction

Where β_5 is the base estimate relating to biscuits and β_6, β_7 and β_8 relate to the offsets for the other three categories in turn.

5.4.2.2 BPM Variable construction for the Fixed Effects Combined Model

As with the separate model, the model is constructed where the Informational reinforcement within the lower Utilitarian reinforcement group is treated as a base and the Informational reinforcement within the higher Utilitarian group being an offset to this base as discussed previously with the separate model construction. The added complication here is the need to introduce the multiple categories to the model. Therefore, each variable now has a base category of biscuits and offsets relating to the other three categories. Hence the Informational reinforcement within the lower Utilitarian reinforcement is constructed as follows (Equation 13):

$$
\begin{aligned}
\text{Inf} \mid \text{UT1}_{biscuits} &= \beta_{9biscuits} \\
\text{Inf} \mid \text{UT1}_{fruitjuice} &= \beta_{9biscuits} + \beta_{10\,fruitjuice} \\
\text{Inf} \mid \text{UT1}_{yellowfats} &= \beta_{9biscuits} + \beta_{11yellowfats} \\
\text{Inf} \mid \text{UT1}_{beans} &= \beta_{9biscuits} + \beta_{12beans}
\end{aligned}
$$

Equation 13: BPM Variable construction (lower Utilitarian reinforcement group)

Similarly, the Informational reinforcement within the higher Utilitarian reinforcement group is constructed as follows in Equation 14:

$$\begin{aligned}
\text{Inf} \mid \text{UT2}_{biscuits} &= \beta_{13biscuits} \\
\text{Inf} \mid \text{UT2}_{fruitjuice} &= \beta_{13biscuits} + \beta_{14fruitjuice} \\
\text{Inf} \mid \text{UT2}_{yellowfats} &= \beta_{13biscuits} + \beta_{15yellowfats} \\
\text{Inf} \mid \text{UT2}_{beans} &= \beta_{13biscuits} + \beta_{16beans}
\end{aligned}$$

Equation 14: BPM Variable construction (higher Utilitarian reinforcement group)

The Informational reinforcement within the higher Utilitarian reinforcement group is an offset to the Informational reinforcement within the lower Utilitarian reinforcement group, as discussed within the separate model build. The multiple category model complicates the calculation for obtaining an estimate i.e. in order to gain an estimate for the Informational reinforcement for the higher Utilitarian reinforcement group for the non-biscuit category (let us assume the fruit juice category estimate is desired for example) then the offset for the category and the offset for the higher Utilitarian reinforcement group must be considered. Hence Equation 15:

$$\begin{aligned}
(\text{Inf} \mid \text{UT2}_{fruitjuice}) = \beta_{13}(\text{Inf} \mid \text{UT1}_{biscuits}) + \beta_{13}(\text{Inf} \mid \text{UT1}_{fruitjuice}) + \\
\beta_{13}(\text{Inf} \mid \text{UT2}_{biscuits}) + \beta_{13}(\text{Inf} \mid \text{UT2}_{fruituice})
\end{aligned}$$

Equation 15: BPM Variable construction (Informational/Utilitarian reinforcement combination)

The other categories' parameter estimates follow the same construction.

Therefore, the trade-off is a more complicated mechanism to get to the parameter estimate, however each offset can be compared to ascertain whether differences exist, statistically.

5.4.2.3 Supermarket Own Brand offset of the BPM Variables for the Combined Model

The supermarket own brand and BPM nest of variables is constructed in the same way. The base measure is the Informational reinforcement within the lower Utilitarian reinforcement group for supermarket own brands within the biscuit category. The other three variables represent the offset of the Informational reinforcement within the lower Utilitarian reinforcement group for non-supermarket own brands for each category in turn from the base biscuit category metric, i.e. Equation 16.

$$\begin{aligned}
&\text{Super} \times \text{Inf} \mid \text{UT1}_{biscuits} = \beta_{13biscuits} \\
&\text{Super} \times \text{Inf} \mid \text{UT1}_{fruitjuice} = \beta_{14fruitjuice} + \beta_{13biscuits} \\
&\text{Super} \times \text{Inf} \mid \text{UT1}_{yellowfats} = \beta_{15yellowfats} + \beta_{13biscuits} \\
&\text{Super} \times \text{Inf} \mid \text{UT1}_{beans} = \beta_{16beans} + \beta_{13biscuits}
\end{aligned}$$

Equation 16: Supermarket own brand offset (lower Utilitarian reinforcement group)

This metric for the biscuits category in itself is an offset to the Informational reinforcement within the lower Utilitarian reinforcement group for the biscuit category, i.e. it is the difference that the supermarket own brand makes. Since it is an offset, the inference for the parameter will facilitate the statistical significance the supermarket own brand effect to be established. In order to construct an estimate for the entire effect within the biscuit category, then both must be summed, i.e. Equation 17.

$$\text{Super} \times \text{Inf} \mid \text{UT1}_{biscuits} = \beta_9(\text{Inf} \mid \text{UT1}_{biscuits}) + \beta_{17}(\text{Super} \times \text{Inf} \mid \text{UT1}_{biscuits})$$

Equation 17: Constructing supermarket own brand offset (base category, lower Utilitarian group)

When constructing the parameter for the other categories, the process is more complex as the individual estimates are themselves offsets to the biscuit category, hence these also need to be taken into consideration in the following manner (Equation 18).

$$\begin{aligned}
\text{Super} \times \text{Inf} \mid \text{UT1}_{\text{fruitjuice}} = {} & \beta_9(\text{Inf} \mid \text{UT1}_{biscuits}) + \beta_{10}(\text{Inf} \mid \text{UT1}_{fruitjuice}) \\
& + \beta_{17}(\text{Super} \times \text{Inf} \mid \text{UT1}_{biscuits}) + \beta_{18}(\text{Super} \times \text{Inf} \mid \text{UT1}_{fruitjuice})
\end{aligned}$$

Equation 18: Constructing supermarket own brand offset (offset category, lower Utilitarian group)

The other categories' parameter estimates follow the same construction.

The Informational Reinforcement within the higher Utilitarian reinforcement group for supermarket own branded products is constructed in a similar fashion, substituting the lower Utilitarian reinforcement group for the higher. The base category is biscuits and the other categories are offsets to this.

In order to build the estimate for the Supermarket own brand effect for the Informational reinforcement within the higher Utilitarian reinforcement group, all the elements must be taken into consideration. For the base biscuits category, this can be constructed as follows (Equation 19).

$$\begin{aligned}\text{Super} \times \text{Inf} \mid \text{UT1}_{biscuits} = {} & \beta_9(\text{Inf} \mid \text{UT1}_{biscuits}) + \beta_{13}(\text{Inf} \mid \text{UT2}_{biscuits}) \\ & + \beta_{17}(\text{Super} \times \text{Inf} \mid \text{UT1}_{biscuits}) + \beta_{21}(\text{Super} \times \text{Inf} \mid \text{UT2}_{biscuits})\end{aligned}$$

Equation 19: Constructing supermarket own brand offset (base category, higher Utilitarian group)

For the other categories, the offsets versus the base biscuit category must also be built into the parameter estimate (Equation 20).

$$\begin{aligned}\text{Super} \times \text{Inf} \mid \text{UT2}_{fruitjuice} = {} & \beta_9(\text{Inf} \mid \text{UT1}_{biscuits}) + \beta_{10}(\text{Inf} \mid \text{UT1}_{fruitjuice}) \\ & + \beta_{13}(\text{Inf} \mid \text{UT2}_{biscuits}) + \beta_{14}(\text{Inf} \mid \text{UT2}_{fruitjuice}) \\ & + \beta_{17}(\text{Super} \times \text{Inf} \mid \text{UT1}_{biscuits}) + \beta_{18}(\text{Super} \times \text{Inf} \mid \text{UT1}_{fruitjuice}) \\ & + \beta_{21}(\text{Super} \times \text{Inf} \mid \text{UT2}_{biscuits}) + \beta_{22}(\text{Super} \times \text{Inf} \mid \text{UT2}_{fruitjuice})\end{aligned}$$

Equation 20: Constructing supermarket own brand offset (offset category, higher Utilitarian group)

With the other categories being built in a similar way to the fruit juice category.

5.4.2.4 Christmas week effect and the BPM Variables

The Christmas week effect is set up in a similar way to the structure of the supermarket own brand, in that the parameter is seen as an offset to represent the change observed within the Christmas week versus any other average week. The structure of the parameters in the model is therefore of the same nature as that of the supermarket own brand indicator. The BPM variables are structured in the same manner, therefore the Informational reinforcement variable is estimated within the lower Utilitarian reinforcement group. The biscuit category is used as the base category and the other categories are offsets to it, allowing the inference of the statistical differences between the parameters to be assessed versus the base category. These are expressed as such (Equation 21).

$$\begin{aligned}
&\text{Christmas } \times \text{Inf } | \text{UT1}_{biscuits} = \beta_{23biscuits} \\
&\text{Christmas } \times \text{Inf } | \text{UT1}_{fruitjuice} = \beta_{24fruitjuice} + \beta_{23biscuits} \\
&\text{Christmas } \times \text{Inf } | \text{UT1}_{yellowfats} = \beta_{25yellowfats} + \beta_{23biscuits} \\
&\text{Christmasr } \times \text{Inf } | \text{UT1}_{beans} = \beta_{26beans} + \beta_{23biscuits}
\end{aligned}$$

Equation 21: Christmas week offset (lower Utilitarian reinforcement group)

In order to estimate the Christmas Informational reinforcement for the biscuit category, it needs to be added to the Informational reinforcement as the variable is an offset for the Christmas week. Therefore, the metrics for the biscuit category can be constructed as follows (Equation 22):

$$\text{Christmas } \times \text{Inf } | \text{UT1}_{biscuits} = \beta_9(\text{Inf } | \text{UT1}_{biscuits}) + \beta_{23}(\text{Christmas } \times \text{Inf } | \text{UT1}_{biscuits})$$

Equation 22: Constructing supermarket own brand offset (base category, lower Utilitarian group)

The other categories are themselves offsets as with the supermarket own brand indicator) and their values can be derived as such (Equation 23):

$$\begin{aligned}
\text{Christmas } \times \text{Inf } | \text{UT1}_{\text{fruitjuice}} = {} & \beta_9(\text{Inf } | \text{UT1}_{biscuits}) + \beta_{10}(\text{Inf } | \text{UT1}_{fruitjuice}) \\
& + \beta_{17}(\text{Christmas } \times \text{Inf } | \text{UT1}_{biscuits}) + \beta_{18}(\text{Christmas } \times \text{Inf } | \text{UT1}_{fruitjuice})
\end{aligned}$$

Equation 23: Constructing supermarket own brand offset (offset category, lower Utilitarian group)

The other two categories are derived in similar fashion, substituting the relevant category β_i for the category in question.

The final focal parameter is the Christmas effect on the Informational reinforcement within the higher Utilitarian informational reinforcement group. The biscuit category forms the base category for the parameter of Informational reinforcement within the higher Utilitarian reinforcement group and the other categories are an offset to this (Equation 24).

$$\begin{aligned}
&\text{Christmas } \times \text{Inf } | \text{UT2}_{biscuits} = \beta_{27biscuits} \\
&\text{Christmas } \times \text{Inf } | \text{UT2}_{fruitjuice} = \beta_{28fruitjuice} + \beta_{27biscuits} \\
&\text{Christmas } \times \text{Inf } | \text{UT2}_{yellowfats} = \beta_{29yellowfats} + \beta_{27biscuits} \\
&\text{Christmasr } \times \text{Inf } | \text{UT2}_{beans} = \beta_{30beans} + \beta_{27biscuits}
\end{aligned}$$

Equation 24: Constructing Christmas week offset (offset category, higher Utilitarian group)

This is in itself an offset to the Christmas effect within lower Utilitarian reinforcement group. In order to construct the estimate for the total effect of the Christmas effect for the Informational reinforcement within the higher Utilitarian group, it must be taken into account the totality of the base and offset variables. For the biscuit category, this is constructed as follows:

$$\begin{aligned}\text{Christmas} \times \text{Inf} \mid \text{UT2}_{biscuits} = {} & \beta_9(\text{Inf} \mid \text{UT1}_{biscuits}) + \beta_{13}(\text{Inf} \mid \text{UT2}_{biscuits}) \\ & + \beta_{23}(\text{Christmas} \times \text{Inf} \mid \text{UT1}_{biscuits}) + \beta_{27}(\text{Christmas} \times \text{Inf} \mid \text{UT2}_{biscuits})\end{aligned}$$

Equation 25: Constructing Christmas week offset (high UT group, base category, higher Utilitarian group)

For the other categories, the offsets versus the base biscuit category must also be built into the parameter estimate.

$$\begin{aligned}\text{Christmas} \times \text{Inf} \mid \text{UT2}_{fruitjuice} = {} & \beta_9(\text{Inf} \mid \text{UT1}_{biscuits}) + \beta_{10}(\text{Inf} \mid \text{UT1}_{fruitjuice}) \\ & + \beta_{13}(\text{Inf} \mid \text{UT2}_{biscuits}) + \beta_{14}(\text{Inf} \mid \text{UT2}_{fruitjuice}) \\ & + \beta_{23}(\text{Christmas} \times \text{Inf} \mid \text{UT1}_{biscuits}) + \beta_{24}(\text{Christmas} \times \text{Inf} \mid \text{UT1}_{fruitjuice}) \\ & + \beta_{27}(\text{Christmas} \times \text{Inf} \mid \text{UT2}_{biscuits}) + \beta_{28}(\text{Christmas} \times \text{Inf} \mid \text{UT2}_{fruitjuice})\end{aligned}$$

Equation 26: Constructing Christmas week offset (high UT group, offset category, higher Utilitarian group)

The other categories are built in a similar way to the fruit juice category.

5.4.2.5 Non Focal variables

The non-focal variables are also built as a base and offset structure. Note the non-focal variables are built as two separate structures within each of the four categories. There is the brand flavour/variant and also the pack size. These are independent across category as it makes little sense on an interpretation level to try and combine these across category given the difference in frequency of purchase, size of pack and size of serving for each category. However, there is scope to explore this in further study.
Hence the base flavour and base size for each category will be omitted and the other variances will become offset to these.

The following (Equation 27) shows the functional form structure of the fixed effects model. The following figure shows the additional hierarchical element to represent the hierarchical intercept model (Equation 28).

$$
\begin{aligned}
LN(Volume_j) = \beta_0 \\
&+ \beta_1(\text{offset for fj}) \\
&+ \beta_2(\text{offset for yf}) \\
&+ \beta_3(\text{offset for beans}) \\
&+ \beta_4 LN(\Pr ice_j) \\
&+ \beta_5 LN(\Pr ice_j \text{offset for fj}) \\
&+ \beta_6 LN(\Pr ice_j \text{offset for yf }) \\
&+ \beta_7 LN(\Pr ice_j \text{offset for beans}) \\
&+ \beta_8 * Informational_j * Utilitarian_j \\
&+ \beta_9 * (Informational_j * Utilitarian_j)\text{offset for fj} \\
&+ \beta_{10} * (Informational_j * Utilitarian_j)\text{offset for yf} \\
&+ \beta_{11} * (Informational_j * Utilitarian_j)\text{offset for beans} \\
&+ \beta_{12} Informational_j * Utilitarian_{jGroup2} \\
&+ \beta_{13}(Informational_j * Utilitarian_{jGroup2})\text{offset for fj} \\
&+ \beta_{14}(Informational_j * Utilitarian_{jGroup2})\text{offset for yf} \\
&+ \beta_{15}(Informational_j * Utilitarian_{jGroup2})\text{offset for beans} \\
&+ \beta_{16} Supermarket * Informational_j \\
&+ \beta_{17}(Supermarket * Informational_j)\text{offset for fj} \\
&+ \beta_{18}(Supermarket * Informational_j)\text{offset for yf} \\
&+ \beta_{19}(Supermarket * Informational_j)\text{offset for beans} \\
&+ \beta_{20} Supermarket * Informational_j * Utilitarian_{jGroup2} \\
&+ \beta_{21}(Supermarket * Informational_j * Utilitarian_{jGroup2})\text{offset for fj} \\
&+ \beta_{22}(Supermarket * Informational_j * Utilitarian_{jGroup2})\text{offset for yf} \\
&+ \beta_{23}(Supermarket * Informational_j * Utilitarian_{jGroup2})\text{offset for beans} \\
&+ \beta_{24} Christmas * Informational_j \\
&+ \beta_{25}(Christmas * Informational_j)\text{offset for fj} \\
&+ \beta_{26}(Christmas * Informational_j)\text{offset for yf} \\
&+ \beta_{27}(Christmas * Informational_j)\text{offset for beans} \\
&+ \beta_{28} Christmas * Informational_j * Utilitarian_{jGroup2} \\
&+ \beta_{29}(Christmas * Informational_j * Utilitarian_{jGroup2})\text{offset for fj} \\
&+ \beta_{30}(Christmas * Informational_j * Utilitarian_{jGroup2})\text{offset for fj} \\
&+ \beta_{31}(Christmas * Informational_j * Utilitarian_{jGroup2})\text{offset for fj} \\
&+ \ldots
\end{aligned}
$$

$$
\begin{aligned}
LN(Volume_j) = \ldots.\\
&+\beta_{32}\text{Chocolate coated (bis)}\\
&+\beta_{33}\text{Plain Sweet (bis)}\\
&+\beta_{34}\text{Filled (bis)}\\
&+\beta_{35}\text{Non Sweet (bis)}\\
&+\text{Countlines (bis)} && \textit{base}\\
&+\beta_{36}\text{Size2 - 5 (bis)}\\
&+\beta_{37}\text{Size6 - 7 (bis)}\\
&+\beta_{38}\text{Size8 - 11 (bis)}\\
&+\beta_{39}\text{Size12 + (bis)}\\
&+\beta_{40}\text{Size packs (bis)}\\
&+\text{Size1s (bis)} && \textit{base}\\
&+\beta_{41}\text{Other fruit (fj)}\\
&+\beta_{42}\text{Breakfast (fj)}\\
&+\beta_{43}\text{Grape (fj)}\\
&+\beta_{44}\text{Grapefruit (fj)}\\
&+\beta_{45}\text{Mixed (fj)}\\
&+\beta_{46}\text{Orange (fj)}\\
&+\beta_{47}\text{Pineapple (fj)}\\
&+\beta_{48}\text{Tomato (fj)}\\
&+\beta_{49}\text{Vegetable (fj)}\\
&+\beta_{50}\text{Vitamin (fj)}\\
&+\text{Apple (fj)} && \textit{base}\\
&+\beta_{51}\text{Size2 - 5 (fj)}\\
&+\beta_{52}\text{Size6 + (fj)}\\
&+\text{Size1s (fj)} && \textit{base}\\
&+\beta_{53}\text{Butter (yf)}\\
&+\beta_{54}\text{Margarine (yf)}\\
&+\beta_{55}\text{Low Reduced (yf)}\\
&+\text{Blended Spreads (yf)} && \textit{base}\\
&+\beta_{56}\text{Size2 + (fj)}\\
&+\text{Size1s (fj)} && \textit{base}\\
&+\beta_{57}\text{Beans Plus (beans)}\\
&+\beta_{58}\text{Tomato (beans)}\\
&+\beta_{59}\text{Healthy (beans)}\\
&+\beta_{60}\text{Flavours (beans)}\\
&+\text{Beans Only (beans)} && \textit{base}\\
&+\beta_{61}\text{Size4 - 12 (beans)}\\
&+\beta\ \text{Size1 - 2 (beans)} && \textit{base}\\
&+\varepsilon_{\mathrm{j}}\\
where\, \varepsilon_{\mathrm{j}} \sim N(0,\sigma^2)
\end{aligned}
$$

Equation 27: Fixed effects functional form (non-hierarchical)

$$
\begin{aligned}
&\beta_0 = U_0 + \nu_0\\
&\nu_0 \sim N(0,\sigma^2)
\end{aligned}
$$

Equation 28: Hierarchical intercept element

Also, the prior distributions are defined. The prior distributions are vague in nature and hence defined as per Equation 29.

$$\beta_i \sim N(0, 0.001)$$
$$\text{for } i = 1,2,\ldots,61$$

$$\sigma \sim Gamma(0.0001, 0.0001)$$

Equation 29: Prior distributions for the fixed effect non-hierarchical model

The prior distributions (also vague) for the hierarchical model is defined as per Equation 30.

$$\beta_i \sim N(0, 0.001)$$
$$\text{for } i = 1,2,\ldots,61$$

$$\upsilon_0[k] \sim N(0, 0.001)$$
$$\text{for } k = 1,\ 2,\ldots,1689$$

$$\sigma \sim Gamma(0.0001, 0.0001)$$
$$\nu_0 \sim Gamma(0.0001, 0.0001)$$

Equation 30: Prior distributions for the fixed effect non-hierarchical model

As with the separate analysis, each non-focal variable is relevant only within its category. Hence a value of 0 is attributed to a non-focal variable outside of its category of interest. Hence the following Fig 60 shows how each of these variables are constructed.

Modelling Variables ====>

Biscuits					Fruit Juice											Yellow Fats				Beans				
Chocolate Coated	Plain Sweet	Filled	Non Sweet	Countlines	Other fruit	Breakfast	Grape	Grapefruit	Mixed	Orange	Pineapple	Tomato	Vegetable	Vitamin	Apple	Butter	Margarine	Low Reduced	Blended spreads	Beans Plus	Tomato	Healthy	Flavours	Beans Only
1	0	0	0	base	0	0	0	0	0	0	0	0	0	0	base	0	0	0	base	0	0	0	0	base
0	1	0	0	base	0	0	0	0	0	0	0	0	0	0	base	0	0	0	base	0	0	0	0	base
0	0	1	0	base	0	0	0	0	0	0	0	0	0	0	base	0	0	0	base	0	0	0	0	base
0	0	0	1	base	0	0	0	0	0	0	0	0	0	0	base	0	0	0	base	0	0	0	0	base
0	0	0	0	base	1	0	0	0	0	0	0	0	0	0	base	0	0	0	base	0	0	0	0	base
0	0	0	0	base	0	1	0	0	0	0	0	0	0	0	base	0	0	0	base	0	0	0	0	base
0	0	0	0	base	0	0	1	0	0	0	0	0	0	0	base	0	0	0	base	0	0	0	0	base
0	0	0	0	base	0	0	0	1	0	0	0	0	0	0	base	0	0	0	base	0	0	0	0	base
0	0	0	0	base	0	0	0	0	1	0	0	0	0	0	base	0	0	0	base	0	0	0	0	base
0	0	0	0	base	0	0	0	0	0	1	0	0	0	0	base	0	0	0	base	0	0	0	0	base
0	0	0	0	base	0	0	0	0	0	0	1	0	0	0	base	0	0	0	base	0	0	0	0	base
0	0	0	0	base	0	0	0	0	0	0	0	1	0	0	base	0	0	0	base	0	0	0	0	base
0	0	0	0	base	0	0	0	0	0	0	0	0	1	0	base	0	0	0	base	0	0	0	0	base
0	0	0	0	base	0	0	0	0	0	0	0	0	0	1	base	0	0	0	base	0	0	0	0	base
0	0	0	0	base	0	0	0	0	0	0	0	0	0	0	base	1	0	0	base	0	0	0	0	base
0	0	0	0	base	0	0	0	0	0	0	0	0	0	0	base	0	1	0	base	0	0	0	0	base
0	0	0	0	base	0	0	0	0	0	0	0	0	0	0	base	0	0	1	base	0	0	0	0	base
0	0	0	0	base	0	0	0	0	0	0	0	0	0	0	base	0	0	0	base	1	0	0	0	base
0	0	0	0	base	0	0	0	0	0	0	0	0	0	0	base	0	0	0	base	0	1	0	0	base
0	0	0	0	base	0	0	0	0	0	0	0	0	0	0	base	0	0	0	base	0	0	1	0	base
0	0	0	0	base	0	0	0	0	0	0	0	0	0	0	base	0	0	0	base	0	0	0	1	base

Figure 60: Structure of non-focal modelling variables

5.4.3 Discussion on Informative Priors for an Offset Model

The models intended for the combined category models exclude the hierarchical model with informative priors which was not the case for the separate category models. The functional form of the offset model for the combined categories can be very complicated. There are offsets for each of the sets of variable but these are also offset to the base category and hence a large number of levels and moving parts. For example, the supermarket own brand effect on the informational reinforcement within the higher Utilitarian reinforcement group for fruit juice categories, relies on a build of coefficients encompassing the Informational reinforcement (base and fruit juice) the offset for both categories for the higher Utilitarian group and further offsets for the supermarket effect of both categories. This complexity, coupled with the lack of past information relating to any prior information on these could lead to a mis-information being applied to the model by the researcher, agreeing with the arguments of (for example) Leamer (1992), Rossi and Allenby, (2003), Gelman (2010). Hence a vague prior distribution is used instead, highlighting both the argument that inclusion of prior distributions brings unnecessary complexity but also how the use of a vague prior can overcome these complexities. When sufficient information is derived from running such offsets, then these can be used as future informative priors of course (O' Hagan, 1994; Duncan et al, 1996).

5.4.4 Interpreting Model Parameter estimates

As discussed earlier, the disadvantage of using offset estimates for the parameter is the need to reconstruct the estimates taking into account the base and offset variants. This text now moves on to discuss how these parameters are reconstructed to form the point estimate and the confidence intervals of each parameter.

5.4.4.1 Reconstructing the point estimates

The offsets can be reconstructed to form a point estimate of each parameter. The offset of each mean estimate is added to the base category to achieve the value of the coefficient for each offset category. For example, the estimate for the beans price elasticity would be as per Equation 31.

$$\beta_{1_{beans}} = \beta_{1_{biscuits}} + \vartheta_{1_{beans}}$$

Equation 31: Point estimate reconstruction

It is also useful to obtain the confidence interval in order to evaluate if the estimate of the parameter is statistically different from zero. The confidence is derived from the confidence interval of the two estimates, i.e. the base and the offset using Equation 32.

$$(\beta_{1_{biscuits}} + \vartheta_{1_{beans}}) \pm \sqrt{\frac{(n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_2^2}{(n_1 + n_2 - 2)}}$$

Equation 32: Confidence interval reconstruction

5.5 Modelling the Data

5.5.1 The Gibbs Sampler

As discussed previously the historic issue with Bayesian inference is the prohibitive nature of calculating the posterior integral for any functional forms other than trivial models. In order to surmount this, the simulation technique of MCMC is employed whereby a sufficiently large number of *iid* draws are made until convergence of the posterior distribution of $p(\theta \mid y)$ is achieved (Rossi and Allenby, 2003). This method is what has been termed Markov Chain Monte Carlo, or MCMC (Robert and Casella, 1999). A suitable algorithm is required to achieve this convergence and the Metropolis-Hastings method has been shown to converge at a geometric rate (Tierney, 1994). The Gibbs sampler (depicted in Equation 33) is one form of Metropolis-Hastings algorithm. Consider the posterior distribution with θ_k elements $(\theta_1,...,\theta_k)$. The Gibbs sampler works by drawing from conditional distributions of the posterior by cycling through each parameter, one at a time whilst maintaining the other parameters constant in the following fashion.

$$\begin{aligned}
&p(\theta_{r,1} \mid \theta_{r-1,2}, \theta_{r-1,3}, ..., \theta_{r-1,k}, y) \\
&p(\theta_{r,2} \mid \theta_{r,1}, \theta_{r-1,3}, ..., \theta_{r-1,k}, y) \\
&... \\
&p(\theta_{r,k} \mid \theta_{r,1}, \theta_{r,2}, ..., \theta_{r,k-1}, y)
\end{aligned}$$

Equation 33: MCMC Gibbs Sampler

This continues until the joint posterior distribution converges. Inference can then be derived for each of the parameters $(\theta_1,...,\theta_k)$ by calculating the estimate for the parameter from the iterations of the converged chain.

The modelling process is conducted through the Rjags package, within the R software system. The Rjags package calls on the JAGS (Just Another Gibbs Sampler) software package and brings its functionality within the R environment (see Plummer (2003) for details on the JAGS package). The JAGS R package uses the Gibbs sampler to generate the model's MCMC, and the CODA package within R offers a suitable means of calculating this Bayesian inference of the parameters (see Finley (2013)).

5.5.2 Convergence Criteria

There is no mechanism whereby the Gibbs sampler "knows" it has converged and the researcher must ensure convergence is achieved before inference can be calculated.

The Bayesian model uses MCMC to calculate the estimate and hence produces a chain of evolving estimates of the parameter value, starting at an arbitrary initial value and through the Gibbs sampler, arrives at a converged estimate of the value of the parameter. Each draw from the chain is autocorrelated, though the laws of large numbers allows the estimations to be inferenced when the chain converges (Rossi and Allenby, 1999). When the chains reach convergence, it is said they resemble "hairy caterpillars" which is a random noise around a stationary value of the estimate. This allows a visual means of assessing if the model has been run with sufficient number of draws to arrive at the estimate.

As well as the visual inspection of the MCMC to ascertain convergence, Gelman and Rubin (1992) offers a diagnostic which helps determine if convergence has been achieved. In essence the statistic measures the difference in variance between chains versus within chain. A value close to 1 indicates convergence. A rule of thumb states a value of less than 1.1 is sufficient to indicate the parameter has converged. The statistic can be calculated within the CODA Bayesian diagnostic package which can be called through the R environment (Finley, 2013).

5.5.3 Number of Chains

More than one chain can be run to estimate the coefficients and Rossi and Allenby (2003) state that this can often be beneficial as convergence can be seen by the intermingle of both chains. Gelman and Rubin (1992) also suggest multiple chains when running MCMC estimation. Each chain is independent of each other and will converge to the same estimate of the parameter given sufficient number of draws. This convergence to the same estimate also offers the further reassurance the estimate has indeed converged. An example of a converged MCMC "hairy caterpillar" plot with two chains is shown in Fig 61 whereby the red and blue colours represent two independent chains.

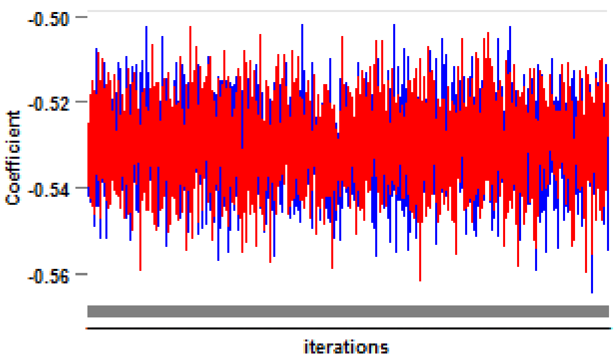

Figure 61: Converged MCMC plot with two chains

Therefore, when the model is constructed, the Gibbs sampler is run using two independent chains for each parameter estimate. The initial values for each parameter in the two chains will be drawn at random from the prior distribution of the parameter and hence each chain will start from a different initial value, offering a further degree of reassurance of the converged of the parameter estimate.

5.5.4 Estimate of the parameter

The parameter is estimated by taking an average of the draws within and then across both chains. Given the initial value could be significantly different from the converged value of the parameter, it is important to base the estimate of the parameter on the average of the converged values rather than the average of the entire chains. To ensure this a "burn in" sample of draws is required and hence the inference is estimated only from the converged draws of the chains. The burn in is set at 4,000 iterations per chain. A further 2,000 iterations per chain are used as the basis of the parameter estimate. There is no rule as to the number of burn in draws and hence it is important to ensure the convergence criteria are checked for all parameter estimates.

5.5.5 Initial Values

Before the model can be initialised, the Gibbs sampler must be given an initial starting value for each chain and each parameter in order to have a base in which to start the Gibbs sampling algorithm. The starting value can be given to the model if an appropriate estimate is known. Otherwise the Rjags package will randomly select a value from the prior distribution assigned (Plummer, 2003). For this study, the latter option is taken and the initial values are sampled from the prior distribution which will result in different starting values for each chain of each parameter, to better ascertain if convergence has been reached (Rossi and Allenby, 2003). The choice of initial value will not make an impact on the parameter estimate, given the inference is taken post burn in, though could make a difference to the number of draws required to reach convergence.

5.5.6 Model File

The Rjags package reads an external data file containing the model functional form, including the prior distribution specification. This is stored as a text file and is called by the body of the model through Rjags.

5.5.7 Generated Statistics

The combination of the MCMC post burn-in iterations are run using the Gibbs sampler resulting in the posterior distribution estimate of each parameter together with its inference. The CODA package within R is a popular means of calculating this inference (Finley, 2013). The posterior distribution is normally distributed and a chart is displayed for each variable using the GGPLOT package within R. Given the Bayesian inference, a 95% confidence interval of the posterior distribution can be observed directly from the MCMC output. A boxplot is also produced through GGPLOT which helps to visualise the difference between comparable parameter estimates. This is helpful when visualising differences or similarities between parameters given various functional forms.

The inference measures are also displayed for each parameter in the form of a point estimate and its standard error. Unlike the frequentist environment, there is no hypothesis test to understand the statistical significance of the point estimate. Instead, the paradigm takes advantage of the fact that the posterior distribution is the probability of the parameter given the data $P(\beta \mid \theta)$ and hence a 95% posterior confidence interval can be calculated for the mean in the usual manner i.e. $\beta \pm 1.96\sigma$. If the 2.5% and 97.5% estimates of the confidence do not straddle zero, then there is at least a 95% probability the value of the parameter is non-zero as illustrated in Fig 62.

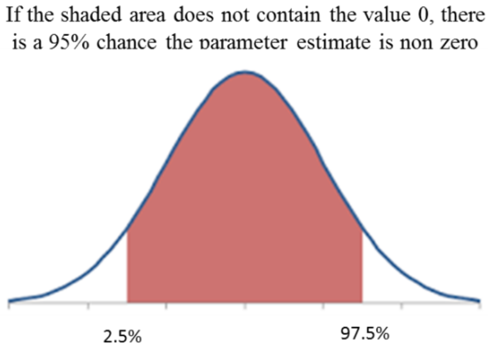

Figure 62: Bayesian posterior confidence interval chart

This measure is used to understand whether the parameter is contributing to the model (if the posterior confidence interval does not straddle zero) or whether the parameter is redundant within the model (i.e. the posterior confidence interval *does* straddle zero). The Bayesian inference allows transparency of course in that it can be easily deduced from the confidence interval whether the degree of confidence the researcher may have as to whether the parameter is "*just*" included/excluded from the interval or whether it is "*some distance*" from the upper/lower confidence interval extremity.

For this study, a combination of Bayesian and frequentist measures will be used to understand the inference of the parameters, given the discussion within the literature review. Fig 63 gives an illustration of the structure of the parameter inference and how these statistics can be interpreted. An indication of whether these are Bayesian or frequentist is also offered.

	(1)	(2)	(3)	(4)	(5)	(6)
	Beta (SE posterior)	Bayes CI		t	sig	
Constant	4.489 (0.0096)	4.47, 4.507	^	467.6	0.000	**
Log Price	-0.701 (0.004)	-0.709, -0.693	^	-175.3	0.000	*
etc.	…					

Figure 63: Parameter interpretation

The estimates and diagnostics of the model parameters are calculated and displayed in tables with headings similar to the one shown in Fig 63.

Each of the metrics of Fig 63 are outlined as follows

(1) Point estimate of the parameter (and its standard error) calculated from the posterior distribution of the MCMC.
(2) The 95% Bayesian posterior confidence interval of the parameter.
(3) The symbol ^ denotes the interval does not straddle zero (and hence means the parameter has at least a 95% probability it is contributing to the model fit). Lack of ^ denotes the interval does straddle zero.
(4) The frequentist t-statistic denotes the ratio of the parameter estimate and its standard error.
(5) The frequentist statistical two-tailed significance level associated with the computed t-statistic.
(6) Indication of the statistical significance with * denoting significance at 10% level and ** denoting significance at the 5% level (two tailed). Lack of stars indicate the level of statistical significance is >10%.

5.6 Assessing the Model Criteria

The Deviance Information Criteria (DIC) is the combination of the "goodness of fit" – "complexity" (Spiegelhalter et al 2002), where the complexity takes into account the number of parameters used in the model (similar in concept to the R-squared (adjusted) frequentist measure). The DIC is a generalisation of AIC and particularly useful when the posterior distribution has been generated from an MCMC estimation approach. The DIC has been a favoured approach of model assessment especially since its incorporation into Bayesian analysis software such as BUGS (see Spiegelhalter *et al.*, 2002 for details). Despite its criticisms there is little alternative currently (Gelman *et al.*, 2013). Taken from Spiegelhalter et al (2002), the fit (shown in Equation 29) is the deviance of the likelihood $p(y\,|\,\theta)$ is defined as in Equation 34.

$$D(\theta) = -2\log L(data\,|\,\theta)$$

Equation 34: Deviance of the likelihood

The "complexity" is defined as the posterior mean deviance plus the deviance of each of the means of each parameter and hence a form of penalty imposed for a more complex model, i.e. Equation 35.

$$\begin{aligned} p_D &= E_{\theta|y}[D] - D(E_{\theta|y}[\theta]) \\ &= \overline{\mathrm{D}} - D(\bar{\theta}) \end{aligned}$$

Equation 35: Model penalty

The DIC shown in Equation 36 is hence constructed in similar means to the AIC as in

$$\begin{aligned} DIC &= D(\bar{\theta}) + 2p_D \\ &= \overline{\mathrm{D}} + p_D \end{aligned}$$

Equation 36: DIC

The smaller the DIC the better the models supports the underlying data.

5.6.1 R-squared (adjusted)

Within the modelling process, the hierarchical and non-hierarchical models will be compared for their explanatory power of the data. Given the differing number of parameters in the models (the hierarchical model will always have a greater number of parameters), the R squared (adjusted) statistic will be used to compare models, given the R squared measure will always increase with more numerous parameters (Field *et al.*, 2012). The R squared (adjusted) attempts to take account of the different number of parameters in each model and adjusts the measure to account for the greater number of parameters within one model compared to the next. Thus, the measure will only increase if the additional parameters are contributing to the model more than can be expected by chance alone (Field et al, 2012). The

R squared adjusts the underlying R squared through the following means, shown in Equation 37, where *n* relates to the number of observations and *p* to the number of parameters.

$$R^2(adj) = 1 - (1 - R^2)\left(\frac{n-1}{p-1}\right)$$

Equation 37: R-squared adjusted

This R squared (adjusted) measure will be used as one means of criticising the model given the benefits discussed earlier of using frequentist methods to help assess the model critique of Bayesian models.

5.6.2 MAPE

The Mean Average Percentage Error is a statistic diagnostic statistic which expresses the average percent difference between the actual and modelled values of a series. The statistic is calculated as shown in Equation 38, where *A* indicates actual values and *M* indicates modelled values.

$$MAPE(\%) = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{A_i - M_i}{A_i}\right|$$

Equation 38: MAPE

Examples of the use of MAPE can be found (e.g. Yang *et al.*, 2006; Maddena and Tanb, 2007; Xu *et al.*, 2010)

5.6.3 Variance Partition Coefficient

Given the hierarchical nature of the model, the variance will be partitioned into two parts, namely the variance between household and the variance between purchaser (Browne and Rasbash, 2004). Let the variance between household be defined as σ_u^2 and the variance between purchasers defined as σ_e^2 then the variance parturition coefficient (Equation 39), which can be expressed as a percentage is defined as

$$VPC = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$$

Equation 39: Variance partition coefficient

The greater the value of the Variance Partition Coefficient (VPC), the larger proportion of the variance is accounted for by the hierarchical structure of the data. A value of zero would indicate the hierarchical structure accounting for zero variance within the model above and beyond the variance accounted for by the non-hierarchical term. In this study, the non-hierarchical variance is that accounted for by variance between purchases and the hierarchical variance accounted for by the variance between household.

In order to test whether the VPC is statistically different from zero, the value of the coefficient of the σ_u^2 can be tested using the frequentist t-test. A null hypothesis of the coefficient is zero is established. A t-statistic is calculated by the ratio of coefficient of the σ_u^2 parameter with its standard error. The value is compared to the t-distribution critical value using a 95% level of significance and null hypothesis rejected if the t-statistic exceeds the critical value. Rejecting the null hypothesis would mean the VPC is statistically different from zero.

The output from the models will display these measures associated with the variance partition coefficient, namely the

Variance (between purchases)
Variance (between households)
Between household t-statistic (with its significance in brackets)
Variance Partition Coefficient

5.6.4 Running the Models (Summary)

The study proceeds by building and running the models discussed in this section. First models will be built separately and run for each of the four categories, namely biscuits, fruit juice, yellow fats and beans. For each category, three functional forms of the model will be run, namely non-hierarchical, hierarchical with vague priors and hierarchical with informative priors. MCMC estimation using the Gibbs sampler will be used to evaluate the Bayesian posterior estimate of each parameter. Every model variant will be run with two chains and a burn in of 4,000 iterations per chain and then a further 2,000 iterations per chain for parameter estimation. The Rjags package is used in each case. Model diagnostics will be calculated and discussed, which will encompass the DIC and parameter convergence from a Bayesian perspective, the R-squared (adjusted) and MAPE from a frequentist perspective and also the variance partition coefficient for the hierarchical models. Results will be discussed and compared as to how the estimates vary between the three model variants within category and also how the estimates compare between categories.

Second, one combined category model for the four categories will be built. This model will consist of four functional forms. The first two will be a pooled non-hierarchical model and a pooled hierarchical model, both with vague priors. The second two models will be the fixed effects model, again with a non-hierarchical and hierarchical model, both with vague priors. As per the separate models, the same model diagnostics will be discussed, both Bayesian and frequentist. Also, the parameter estimates and their Bayesian and frequentist inference of each model will be discussed as well as a comparison between the four functional forms.

Chapter 6: Separate Category Analysis

6.1 Introduction

The next section employs the methodology discussed previously on the four categories of products in turn.

For each category and, where relevant, the cleaned data is used within the analysis. The products have been grouped according to type and pack size as discussed. The variables have been calculated relating to price BMP variables, supermarket own brand indicator variable and the Christmas week dummy variable, as discussed earlier. The price variable is logged in line with past studies (e.g. Oliveria-Castro *et al.*, 2006).

The informational variable is divided into two, the base variable and the offset for the higher utilitarian group. This means the base informational coefficient will represent the lower utilitarian group and the offset combined with the base coefficient will represent the higher group. As discussed, this makes assessing the statistical difference between them more transparent. There is also an offset for the supermarket own brand informational offset to differentiate branded and non-branded products.

The week which contains the Christmas holiday is flagged as a Christmas dummy variable. The models are run using three functional forms, though each utilising Bayesian inference to calculate the parameters. The functional forms comprise of hierarchical with vague prior distributions, hierarchical with informative prior distributions and finally non-hierarchical form. The informative nature of the parameters is as discussed in the methodology chapter. In each case a discussion around the model diagnostics and the parameter estimates is offered discussing to what degree the differing functional forms impacts the model estimates.
A conclusion is offered suggesting the hierarchical models perform in a superior way to the non-hierarchical models in each case in terms of model diagnostics.

6.2 Model Description

The models are as described in the methodology chapter are Bayesian models with functional forms of a hierarchical structure with vague prior distributions, a hierarchical structure with informative prior distributions based on the initial frequentist analysis and a non-hierarchical structure.

The models are run with a burn in of 4,000 iterations and the parameters are monitored over a further 2,000 iterations using two independent chains. The convergence charts of the focal parameters for the three models are shown in Figs 1-12 within the Appendix. It can be seen the charts show convergence has been achieved.

The corresponding density charts of the focal parameters show the posterior distributions of the parameters and will be discussed within the body of the text following. The Gaussian nature of the distributions reflects the conjugate nature of the prior as expected, discussed in more detail in the methodology chapter. The small standard deviation (relative to the estimate) of the estimates also suggest the parameters have converged. All three models are therefore presented as converged models and the diagnostics are now discussed.

6.3 Biscuits Model

6.3.1 Model Diagnostics

The diagnostics of the three models are shown in Table 22. From a Bayesian perspective, the Deviance Information Criteria (DIC) is calculated as the sum of the *mean deviance* and the *penalty* to compensate for the relative complexity of the models. More complex models have a higher penalty. It can be seen from Table 22 the penalty for the hierarchical models (vague and informative) is higher than the non-hierarchical model (1,323 for the hierarchical vague, 1,318 for the hierarchical informative and 18 for the non-hierarchical models respectively). The mean deviance for each in turn is 69,379, 69,988 and 81,152. The DIC calculations are therefore 70,702 (hierarchical vague), 71,306 (hierarchical informative) and 81,170 (non-hierarchical). Therefore, the increased penalty incurred by the hierarchical models compared to the non-hierarchical model is outweighed by the increase in the predictive nature of the model. This suggests the hierarchical models would better predict a replicated data set of the same structure (Spiegelhalter *et al.*, 2002). The difference between the hierarchical models

(>5) suggests the vague model is better representing the data than the informative model (Spiegelhalter *et al.,* 2002).

From the frequentist diagnostics, Table 22 shows the R-squared (adj) figures are 55.863% (hierarchical vague), 55.398% (hierarchical informative) and 45.291% (non-hierarchical) suggesting the hierarchical models are explaining a higher proportion of the variance, having accounted for the additional complexity of the models. This agrees with the DIC results. The Mean Average Percentage Error (MAPE) values in respective order are 6.55%, 5.98% and 5.93% showing similar average absolute deviance for the models, though the hierarchical vague model has a larger MAPE.

The total model variance for the hierarchical models is lower than that of the non-hierarchical models (0.182, 0.184 and 0.221 respectively) suggesting the hierarchical structure is representing more of the variability of the data within the model structure. The coefficients of the hierarchical variance term have high t-values when considering their ratio with their standard errors. This offers sufficient evidence to reject the null hypotheses these values are equal to 0. Additionally, the hierarchical variance partition coefficients are $\frac{\sigma_0^2}{\sigma_0^2+\sigma^2} = 17.582\%$ for the hierarchical vague model and 17.413% for the hierarchical informative model.

Despite all three models seeming adequate representations of the underlying data, these statistics suggest the functional form of the hierarchical models is benefitting the model fit above and beyond that of the non-hierarchical form.

6.3.2 Model Coefficients

The coefficients of the models and their inference are displayed in Table 22 and these will be discussed in turn for each parameter in the next section. First, the convergence of the parameters needs to be assessed. Figs 1-3 in the appendix shows the convergence “hairy caterpillar” type charts for the post burn-in MCMC draws of the focal parameters and their nature suggest convergence has been achieved. Furthermore, the Gelman statistics in Table 21 also indicate convergence of the parameters.

	Hierarchical Vague		Hierarchical Informative		Non Hierarchical	
	Point Estimate	Upper CI	Point Estimate	Upper CI	Point Estimate	Upper CI
Constant	1	1.01	1	1	1	1
Log Price	1	1	1	1	1	1
Informational x Utilitarian Gp1	1	1	1	1	1	1
Informational x Utilitarian Gp2	1	1.01	1	1	1	1
SuperOwn x Informational	1	1	1	1	1	1
SuperOwn x Informational GP2	1	1	1	1	1	1
Chrsitmas	1	1	1	1	1	1
Chrsitmas Ut Gp2	1	1	1	1	1	1
Chocolate Coated	1	1	1	1	1	1
Plain Sweet	1	1.01	1	1	1	1
Filled	1	1.01	1	1	1	1
Non Sweet	1	1.01	1	1	1	1
Size 2-5	1	1	1	1	1	1
Size 6-7	1	1	1	1	1	1
Size 8-11	1	1	1	1	1	1
Size 12+	1	1	1	1	1	1

Table 21: Gelman convergence measures - biscuits

	Non Hierarchical				Hierarchical Vague				Hierarchical Informative			
	Beta (SE posterior)	Bayes CI	t	sig	Beta (SE posterior)	Bayes CI	t	sig	Beta (SE posterior)	Bayes CI	t	sig
Constant	4.489 (0.0096)	4.47, 4.507 ^	467.6	0.000 **	4.541 (0.0105)	4.52, 4.561 ^	432.5	0.000 **	4.390 (0.0094)	4.372, 4.409 ^	467.0	0.000 **
Log Price	-0.701 (0.004)	-0.709, -0.693 ^	-175.3	0.000 **	-0.702 (0.004)	-0.71, -0.695 ^	-175.6	0.000 **	-0.705 (0.0027)	-0.71, -0.7 ^	-261.1	0.000 **
Informational x Utilitarian Gp1	0.027 (0.0035)	0.02, 0.034 ^	7.7	0.000 **	0.033 (0.0034)	0.026, 0.039 ^	9.6	0.000 **	0.055 (0.002)	0.051, 0.059 ^	27.4	0.000 **
Informational x Utilitarian Gp2	0.074 (0.0042)	0.065, 0.082 ^	17.5	0.000 **	0.054 (0.004)	0.046, 0.062 ^	13.5	0.000 **	0.102 (0.0007)	0.101, 0.104 ^	146.1	0.000 **
SuperOwn x Informational	0.008 (0.0036)	0.001, 0.015 ^	2.3	0.030 *	-0.001 (0.0035)	-0.008, 0.005	-0.4	0.368	0.010 (0.0023)	0.005, 0.014 ^	4.1	0.000 **
SuperOwn x Informational GP2	-0.093 (0.005)	-0.102, -0.083 ^	-18.5	0.000 **	-0.081 (0.0049)	-0.09, -0.071 ^	-16.6	0.000 **	-0.061 (0.0037)	-0.068, -0.054 ^	-16.5	0.000 **
Chrsitmas	0.058 (0.0292)	0.001, 0.117 ^	2.0	0.055	0.043 (0.0266)	-0.009, 0.094	1.6	0.111	0.027 (0.0186)	-0.009, 0.064	1.5	0.136
Chrsitmas Ut Gp2	0.008 (0.0439)	-0.08, 0.091	0.2	0.393	-0.015 (0.0405)	-0.094, 0.067	-0.4	0.374	0.039 (0.0276)	-0.014, 0.092	1.4	0.145
Chocolate Coated	0.152 (0.0069)	0.139, 0.166 ^	22.1	0.000 **	0.143 (0.0066)	0.13, 0.156 ^	21.6	0.000 **	0.123 (0.0067)	0.11, 0.136 ^	18.4	0.000 **
Plain Sweet	0.160 (0.0093)	0.142, 0.178 ^	17.2	0.000 **	0.123 (0.009)	0.105, 0.14 ^	13.6	0.000 **	0.212 (0.007)	0.198, 0.225 ^	30.3	0.000 **
Filled	-0.011 (0.0085)	-0.028, 0.005	-1.3	0.162	-0.027 (0.0084)	-0.043, -0.01 ^	-3.3	0.002 **	-0.030 (0.0084)	-0.046, -0.013 ^	-3.6	0.001 **
Non Sweet	0.039 (0.0104)	0.019, 0.059 ^	3.7	0.000 **	-0.017 (0.01)	-0.036, 0.003	-1.7	0.099	0.086 (0.0071)	0.072, 0.099 ^	12.1	0.000 **
Countlines	base				base				base			**
Size 2-5	0.207 (0.0083)	0.19, 0.223 ^	24.9	0.000 **	0.200 (0.0079)	0.184, 0.215 ^	25.3	0.000 **	0.204 (0.008)	0.188, 0.22 ^	25.6	0.000 **
Size 6-7	0.086 (0.0072)	0.072, 0.1 ^	12.0	0.000 **	0.101 (0.0067)	0.089, 0.115 ^	15.1	0.000 **	0.124 (0.0069)	0.11, 0.137 ^	17.9	0.000 **
Size 8-11	0.195 (0.0078)	0.179, 0.21 ^	24.9	0.000 **	0.190 (0.0077)	0.175, 0.205 ^	24.7	0.000 **	0.199 (0.0076)	0.184, 0.214 ^	26.2	0.000 **
Size 12+	0.360 (0.0071)	0.347, 0.374 ^	50.7	0.000 **	0.332 (0.0068)	0.318, 0.345 ^	48.8	0.000 **	0.333 (0.0068)	0.32, 0.347 ^	49.0	0.000 **
Size packs	0.590 (0.01)	0.571, 0.61 ^	59.0	0.000 **	0.583 (0.0093)	0.564, 0.6 ^	62.6	0.000 **	0.585 (0.0094)	0.566, 0.603 ^	62.2	0.000 **
Size 1s	base				base				base			
R-Squared (adj)	45.291%				55.863%				55.398%			
Mean Deviance	81,152				69,379				69,988			
Penalty	18.2				1323.0				1318.0			
DIC	81170				70702				71306			
MAPE	5.93%				6.55%				5.98%			
Variance (between purchases)	0.221				0.182				0.184			
Variance (between housholds)					0.039				0.039			
between household t-stat (sig)					23.135(0)				23.458(0)			
Variance Partition Coeficient					17.582%				17.413%			

* significant 5%
** significant 1%
^ 95% Bayesian estimates do not include zero

Table 22: Model diagnostics and inference - biscuits

The parameters of the focal variables are visualised graphically in Fig 64 below, demonstrating the differences between the hierarchical and non-hierarchical estimates.

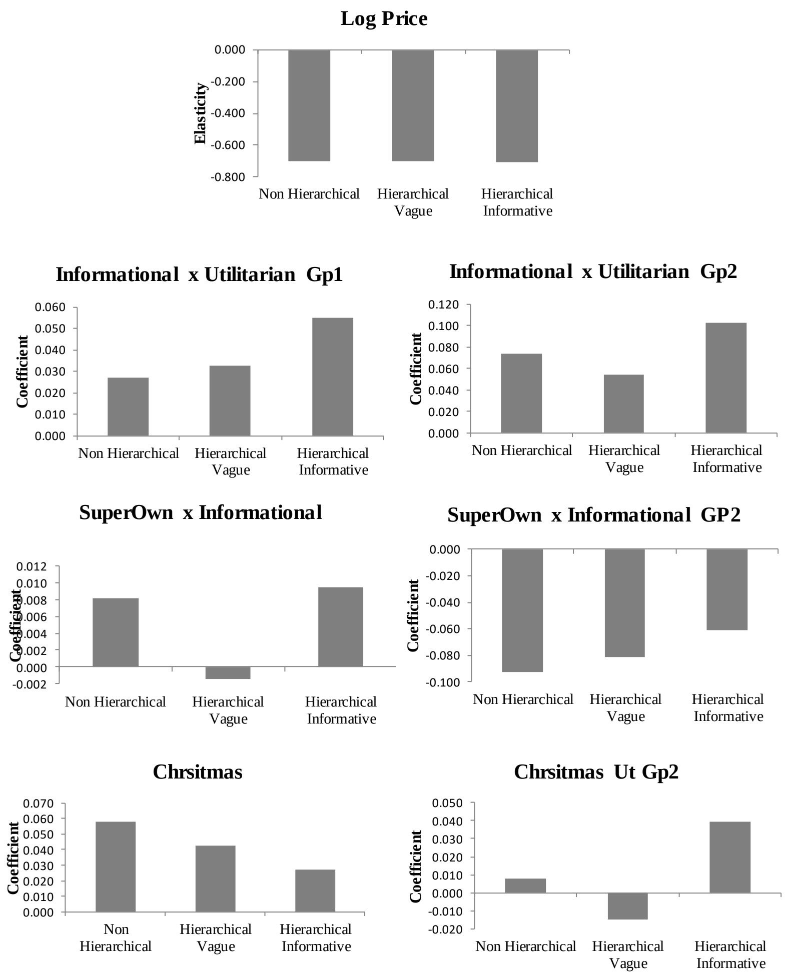

Figure 64: Parameter column charts - biscuits

6.3.2.1 Price Elasticity

Given the log-log model, the coefficient is the *price* elasticity. Fig 65 shows the density plots and box plots of the hierarchical and non-hierarchical models. There is little difference between the elasticity measures of the models. As discussed, the Bayesian nature of the parameter estimate implies the posterior distribution is the probability distribution of the parameter itself and the density plots can be used to understand the shape of the posterior estimates. The point estimates for hierarchical vague, hierarchical informative and non-hierarchical models are -0.702, -0.705 and -0.701 respectively. The 95% Bayesian confidence interval (i.e. between the 2.5% and the 97.5% points on the posterior density plot) for the hierarchical vague model is (-0.71, -0.695) for the hierarchical informative (-0.710, -0.700) and for the non-hierarchical (-0.709, -0.693), none of which include the value zero, hence it can be stated with 95% probability, this parameter is non-zero and hence contributing to the model.

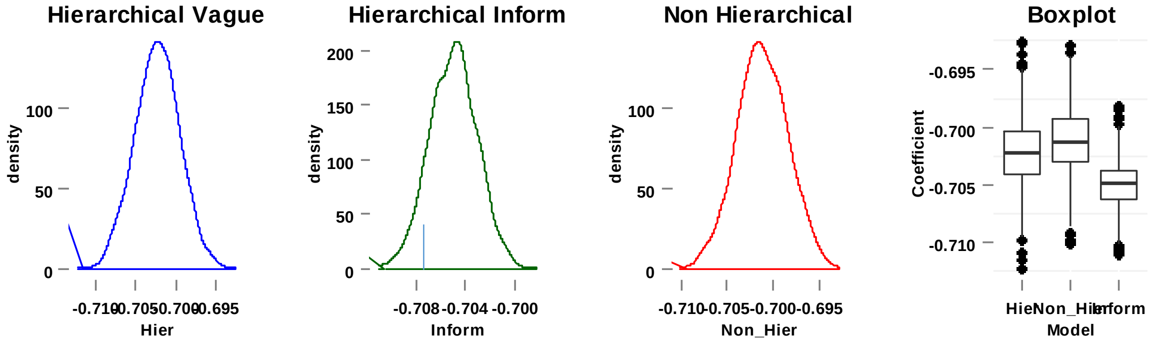

Figure 65: Price coefficients - biscuits

As discussed earlier, studies have benefitted from a range of frequentist and Bayesian inference and hence this approach is employed in this study. From a frequentist perspective, a null hypothesis is constructed the parameter in question is zero. The associated t-statistics of -175.6 for the hierarchical vague and -261.1 for the hierarchical informative and -175.3 for the non-hierarchical which are all statistically significant at $p<0.001$, which leads us to reject the null hypothesis the parameter is equal to zero, offering further evidence the parameter is significantly contributing to the model.

This estimate is aligned with Foxall *et al.*, (2009) who found similar results[1].

6.3.2.2 Informational reinforcement in the lower Utilitarian reinforcement group

The informational variable is the base value and the informational variable for utilitarian group 2 (the higher group) is an offset, hence the base informational coefficient can be interpreted as the value for utilitarian group 1 (the lower utilitarian group). Adding the offset will give the value for utilitarian group 2. The coefficients are transformed to linearity using the transformation shown in Equation 40.

$$\text{linear coef} = e^{\beta} - 1$$

Equation 40: Informational coefficient transformation

.

[1] For non-hierarchical models

Fig 66 shows the posterior distribution density plots and boxplot of the informational variable for the hierarchical and non-hierarchical models.

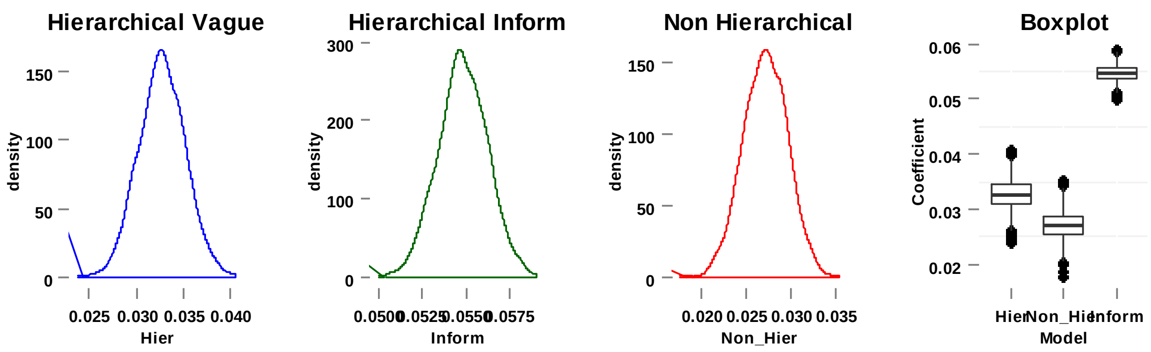

Figure 66: Informational reinforcement (lower Utilitarian group) coefficients - biscuits

The point estimates for the lower utilitarian groups are 0.033, 0.055 and 0.027 for the hierarchical vague, hierarchical informative and non-hierarchical models respectively. In each case, there is very little evidence to suggest this parameter is zero given the Bayesian confidence intervals of (0.026, 0.039) for the hierarchical vague model, (0.051, 0.059) for the hierarchical informative model and (0.020, 0.034) for the non-hierarchical model. None of the models' posterior confidence interval contains the value 0 suggesting the parameters are significant in each case. There is some overlap in the posterior confidence intervals of the non-hierarchical model and the hierarchical vague model. This is due to agreement between the prior distribution of the hierarchical vague model and the likelihood based on the data. Also, the frequentist t-statistic is 9.6, 27.4 and 7.7 respectively, all yielding $p<0.001$, hence strong evidence to suggest the parameter is non-zero in each case. Therefore, the nature of the positive coefficient suggests that larger (volume) brands within the lower utilitarian group are being perceived to have a higher informational benefit than smaller brands, over and above what can be accounted for by price.

6.3.2.3 Informational reinforcement in the higher Utilitarian reinforcement group (offset)

Fig 67 shows the hierarchical and non-hierarchical posterior distribution for the offset informational reinforcement variable for higher utilitarian reinforcement group as a density plot and as a box plot.

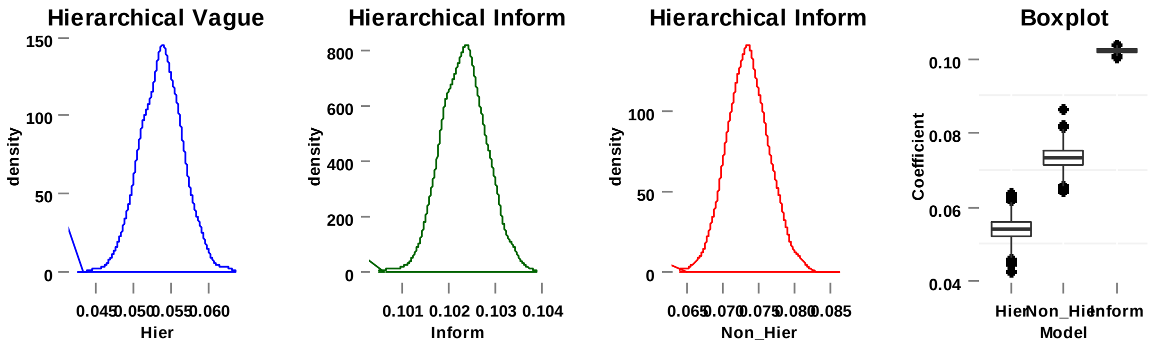

Figure 67: Informational reinforcement (higher Utilitarian group) coefficients - biscuits

The offset value of the coefficient is 0.054 and 0.102 for the hierarchical models in turn and 0.074 for the non-hierarchical model. The Bayesian posterior confidence intervals are (0.046, 0.062), (0.101, 0.1014) and (0.065, 0.082) respectively. Given the intervals do not contain the value zero, there is a 95% probability the parameters are non-zero, hence benefitting model prediction. Also, the frequentist t-statistics for each model are 13.5, 145.1 and 17.5 for the hierarchical vague, hierarchical informative and non-hierarchical models respectively, rejecting the null hypothesis of a zero value parameter. This suggests the informational benefit within the higher utilitarian group is contributing positively to the volume of the category above and beyond the informational benefit within the lower utilitarian group. Despite broad agreement between the models as to the positive nature of the coefficients, all models are suggesting a different magnitude of effect and given the lack of overlap in the posterior confidence intervals, this would imply these are statistically different. Hence the nature of model selected both in terms of structure and prior distribution selection has a differing outcome on the magnitude of the effect of the variable. This is in line with discussions around using the Bayesian paradigm and the importance of prior selection (Rossi and Allenby, 2003).

Combining the results of the two informational variables, it can be seen that, within the BPM structure, having taken account of the price variable, the informational and utilitarian variables are contributing positively to the volume of the biscuit category. The higher the informational values, the higher the volume and the higher utilitarian group is having a higher impact on volume per purchase. This is true for all three model structures.

6.3.2.4 Supermarket own brand x Informational reinforcement in the lower Utilitarian reinforcement group

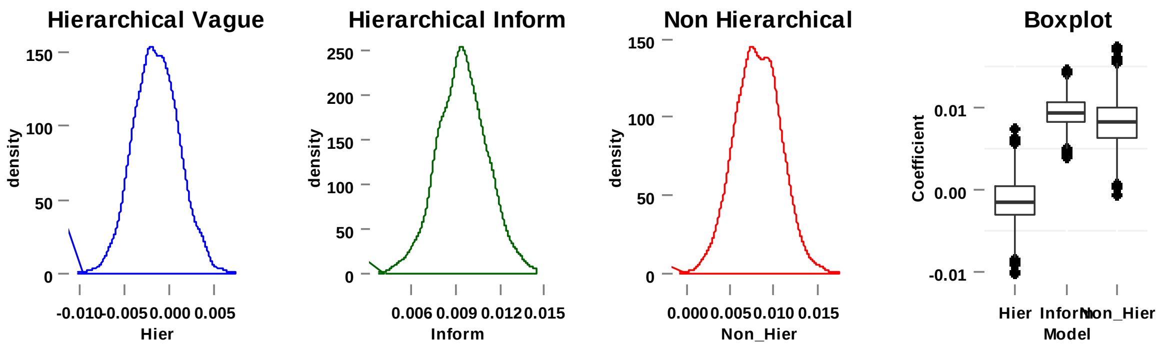

Figure 68: Supermarket own brand x Informational reinforcement (lower Utilitarian group) - biscuits

Fig 68 depicts the density and box plots for the hierarchical (vague and informative) and non-hierarchical models. From Table 22 the coefficient for the hierarchical vague model is -0.001, 0.010 for the hierarchical informative and 0.008 for the non-hierarchical model. The 95% Bayesian confidence intervals for the three models in turn are (-0.008, 0.005), (0.005, 0.014) and (0.004, 0.015), with frequentist t-statistics of -0.4, 4.1 and 2.3 in each case respectively. This demonstrates the hierarchical vague model's parameter is not different from zero, given the Bayesian confidence interval straddles zero and the t-statistics is non-significant (p=0.368). However, the hierarchical informative model and the non-hierarchical model suggest the parameter is positive and statistically significant from both a Bayesian and frequentist standpoint. The informative nature of the hierarchical prior has influenced the result of the hierarchical informative model to have a positive estimate which differs from the hierarchical vague model estimate. This again demonstrates the importance of the prior distribution in model build.

Therefore, differing conclusions as to the nature of the variable and how it may affect sales. The evidence suggests it will be a positive effect or no effect, depending on the model chosen to represent the data.

6.3.2.5 Supermarket own brand x Informational reinforcement in the higher Utilitarian reinforcement group (offset)

Fig 69 shows the density plots and box plots for the parameter estimates of this variable. The point estimates for the three models (in the usual order) are -0.081, -0.061 and -0.093. The

Bayesian posterior confidence intervals for the hierarchical vague and non-hierarchical models overlap, (-0.090, -0.071) and (-0.102, -0.083) suggesting these is agreement between the likelihood and the prior. The confidence interval of the hierarchical informative is higher at (-0.068, -0.054). All intervals do not straddle zero, also the t-statistics are all significant at p<0.001 (values re -16.6, -16.5 and -18.5 respectively). Hence these coefficients are statistically significant in the models. The models suggest the informational reinforcement variable associated with the supermarket own brands within the higher utilitarian reinforcement group are negatively contributing to the volume of the category, above and beyond the effect observed in the lower utilitarian reinforcement group.

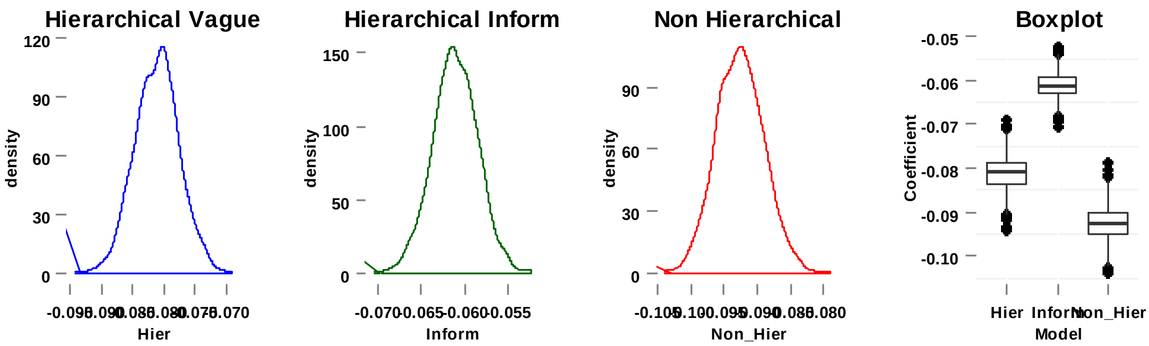

Figure 69: Supermarket own brand x Informational reinforcement (higher Utilitarian group) - biscuits

6.3.2.6 Christmas effect x Informational reinforcement in the lower Utilitarian reinforcement group

From the exploratory data analysis conducted within Chapter 3, the discussion suggests the week containing the Christmas holiday has a noticeably lower volume than other weeks and the inclusion of the dummy variable to test this is discussed in the methodology chapter. Fig 70 shows the usual charts of the inference.

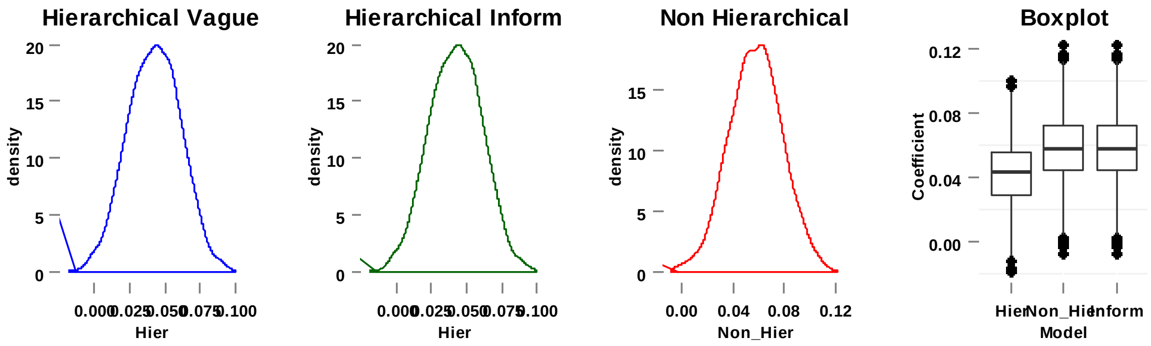

Figure 70: Christmas effect x Informational reinforcement (lower Utilitarian group) - biscuits

The models' estimates of the parameter are 0.043, 0.027 and 0.058 in turn. The Bayesian posterior confidence intervals are (-0.009, 0.094) for the hierarchical vague model, (-0.009, 0.064) for the hierarchical informative model and (0.001, 0.117) for the non-hierarchical model. The respective t-statistics are 1.6 (p=0.111), 1.5 (p=0.136) and 2.0 (p=0.055) for the three modes, suggesting the hierarchical structured models conclude no effect. The non-hierarchical model shows the Bayesian confidence interval does not straddle zero however the frequentist p-value at a strict 95% level is not significant. This does show some disagreement between the paradigms, strictly speaking, though given the proximity of the lower confidence interval to zero and also the marginal significance level (p=0.055). Therefore, a collective viewpoint would be to accept this parameter is having a positive effect on volume purchases.

The variable relates to the volume purchased per transaction and hence despite a lower volume in the period, it would suggest this is due to lower number of transactions rather than lower volume per transaction. This implies the number of transactions (and hence volume) is much lower for this period, however, when transactions are made, the volume bought per transaction is higher. This may be reflective of the deals which are prevalent within the category immediately post-Christmas and consumers are possible making the most of these offers above and beyond what can be explained by the underlying price elasticity measure.

6.3.2.7 Christmas effect x Informational reinforcement in the higher Utilitarian reinforcement group (offset)

Fig 71 shows the density plots and box plots of the posterior distributions of the parameters of the three models.

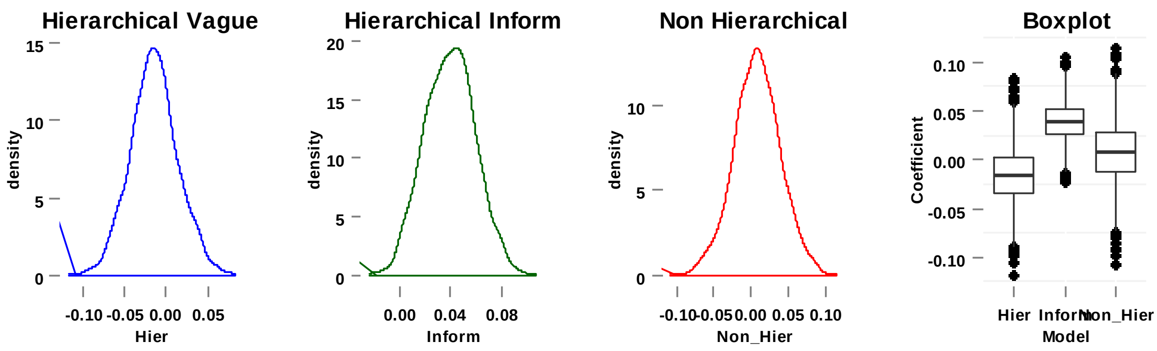

Figure 71: Christmas effect x Informational reinforcement (higher Utilitarian group) - biscuits

The point estimates for the three models are -0.015, 0.039 and 0.008 respectively. The Bayesian posterior confidence intervals for the three models in turn are (-0.094, 0.067), (-0.014, 0.092) and (-0.080, 0.091) and the t-statistics are -0.4, 1.4 and 0.2 in turn, all non-significant at p >=0.145. Therefore, there is no evidence from a Bayesian or frequentist perspective to suggest the Christmas week has an effect on volume sales within the higher utilitarian reinforcement group, above and beyond what can be accounted for by the effect within the lower Utilitarian reinforcement group.

6.3.2.8 Product Characteristic Variables

The product characteristics are dummy variables and the coefficient adjusts the intercept of the model for higher or lower volume levels. The characteristics are biscuit type and pack size. The base biscuit type is countlines and the other variants are offsets to this. The coefficients of the hierarchical models are almost identical. The non-hierarchical model differs with the sign of Non-Sweet being opposite to the hierarchical models. Though the coefficients are small they are statistically significant from a Bayesian and Frequentist perspective. Therefore, the type of biscuits makes a difference to the volume bought per purchase.

The base category for the pack size is the single serve packs. The volume sold in other packs is all statistically significantly higher which makes logical sense given the volume per pack is higher in every case. Consistently across all three models the “pack” type has the higher coefficient which contains the larger weight purchases.

6.4 Fruit Juice

6.4.1 Model diagnostics

Figs 4-6 in the appendix shows the convergence plots for all fixed effects coefficients of the three models. The trace plots of the two chains suggest the parameters have converged given the criteria outlined in the methods chapter. As with the biscuit category, the diagnostics of the models are compared. Table 23 shows the Bayesian diagnostics for the hierarchical vague, hierarchical informative and non-hierarchical models respectively, which in turn show a mean deviance of 24,820, 25,091 and 36,118. The penalty for the models (in the same order) is 846, 841 and 21 resulting in DIC calculation of 25,666, 25,933 and 36,139 for the respective models. Hence the Bayesian diagnostic measures suggest that, despite a larger penalty for a more complicated model structure, the hierarchical models are better constructed to predict a similar data set than the non-hierarchical model (Spiegelhalter *et al.,* 2002). Also, there is evidence to suggest there is a difference in the predictive ability between of the two hierarchical models given the DIC difference between then is >5 (Spiegelhalter *et al.,* 2002) with the vague model with the lower DIC. From the frequentist statistics, the R-squared (adjusted) values for the three models are 55.185%, 54.628% and 20.764% respectively for the hierarchical vague, hierarchical informative and non-hierarchical models, indicating the hierarchical structure is a better representation of the data. There is little difference between the R-squared (adjusted) values of the two hierarchical structures though the vague model is higher. The MAPE values of the models are 4.422%, 4.465% and 6.237% for the three models in turn, suggesting preference for the hierarchical structure as it has a lower mean average percentage error.

The total error variance of the hierarchical models is 0.187 and 0.190 in turn, lower than the equivalent value of 0.318 of the non-hierarchical model. The variance parameter between households of the hierarchical models is 0.144 and 0.145 for the vague and informative models respectively. This coefficient has a t-statistic of 19.118 for the vague model and 19.459 for the informative, both significant at the $p<0.001$ level and hence no evidence to suggest this variance parameter is zero. This results in a variance partition coefficient of 43.409% and 43.334% for the vague and informative models respectively which indicates the hierarchical structure forms a significant proportion of the variance of the models. These combined diagnostic statistics suggest the hierarchical structure is better representing the underlying data, highlighting the importance of the hierarchical structure of the model within this category.

There is very little difference between the vague and informative hierarchical models in terms of model performance.

	Non Hierarchical				Hierarchical Vague				Hierarchical Informative			
	Beta (SE posterior)	Bayes CI	t	sig	Beta (SE posterior)	Bayes CI	t	sig	Beta (SE posterior)	Bayes CI	t	sig
Constant	8.048 (0.0195)	8.011, 8.087 ^	412.718	0.000 **	8.190 (0.0255)	8.139, 8.239 ^	321.157	0.000 **	8.074 (0.0207)	8.034, 8.114 ^	390.063	0.000 **
Log Price	-0.493 (0.0086)	-0.51, -0.476 ^	-57.337	0.000 **	-0.531 (0.0093)	-0.549, -0.513 ^	-57.097	0.000 **	-0.451 (0.0057)	-0.462, -0.439 ^	-79.035	0.000 **
Informational x Utilitarian Gp1	0.200 (0.0086)	0.183, 0.217 ^	23.221	0.000 **	0.128 (0.0086)	0.111, 0.145 ^	14.907	0.000 **	0.060 (0.0058)	0.049, 0.072 ^	10.379	0.000 **
Informational x Utilitarian Gp2	0.039 (0.015)	0.009, 0.068 ^	2.613	0.013 *	0.023 (0.0149)	-0.006, 0.052	1.544	0.121	-0.070 (0.0069)	-0.084, -0.057 ^	-10.174	0.000 **
SuperOwn x Informational	-0.034 (0.0071)	-0.048, -0.02 ^	-4.845	0.000 **	-0.017 (0.007)	-0.03, -0.003 ^	-2.371	0.024 *	0.048 (0.0047)	0.039, 0.057 ^	10.298	0.000 **
SuperOwn x Informational GP2	-0.090 (0.0179)	-0.124, -0.054 ^	-5.006	0.000 **	-0.037 (0.0171)	-0.07, -0.003 ^	-2.170	0.038 *	-0.031 (0.0095)	-0.05, -0.013 ^	-3.274	0.002 **
Chrsitmas	0.005 (0.0392)	-0.073, 0.081	0.125	0.396	0.014 (0.0306)	-0.045, 0.074	0.454	0.360	0.021 (0.0247)	-0.027, 0.07	0.858	0.276
Chrsitmas Ut Gp2	0.074 (0.11)	-0.144, 0.286	0.671	0.319	0.080 (0.0883)	-0.09, 0.258	0.907	0.264	0.071 (0.072)	-0.075, 0.209	0.989	0.245
Other fruit	0.014 (0.0417)	-0.069, 0.094	0.329	0.378	0.048 (0.0371)	-0.024, 0.124 ^	1.302	0.171	0.007 (0.0365)	-0.068, 0.077	0.192	0.392
Breakfast	-0.168 (0.057)	-0.278, -0.055 ^	-2.944	0.005 **	-0.082 (0.0471)	-0.177, 0.009	-1.747	0.087	-0.063 (0.0481)	-0.156, 0.031	-1.301	0.171
Grape	0.107 (0.0246)	0.059, 0.155 ^	4.362	0.000 **	0.153 (0.0218)	0.111, 0.196 ^	7.014	0.000 **	0.150 (0.0217)	0.108, 0.193 ^	6.899	0.000 **
Grapefruit	-0.156 (0.0195)	-0.193, -0.118 ^	-7.974	0.000 **	-0.044 (0.0191)	-0.082, -0.007 ^	-2.319	0.027 *	-0.042 (0.0198)	-0.081, -0.003 ^	-2.131	0.041 *
Mixed	-0.028 (0.0179)	-0.062, 0.008	-1.542	0.122	0.000 (0.0152)	-0.03, 0.03	0.020	0.399	0.043 (0.0154)	0.013, 0.074 ^	2.812	0.008 **
Orange	0.013 (0.0107)	-0.008, 0.034	1.243	0.184	0.030 (0.0099)	0.011, 0.049 ^	3.051	0.004 **	0.036 (0.01)	0.017, 0.056 ^	3.580	0.001 **
Pineapple	-0.275 (0.019)	-0.313, -0.238 ^	-14.447	0.000 **	-0.148 (0.0164)	-0.18, -0.115 ^	-9.006	0.000 **	-0.150 (0.0167)	-0.181, -0.117 ^	-9.006	0.000 **
Tomato	-0.312 (0.029)	-0.368, -0.255 ^	-10.769	0.000 **	-0.240 (0.0284)	-0.295, -0.184 ^	-8.437	0.000 **	-0.223 (0.0287)	-0.281, -0.168 ^	-7.777	0.000 **
Vegetable	0.114 (0.0656)	-0.016, 0.242	1.735	0.089	0.110 (0.0587)	-0.002, 0.227	1.881	0.068	0.095 (0.0575)	-0.018, 0.209	1.645	0.103
Vitamin	-0.139 (0.0949)	-0.325, 0.046	-1.466	0.136	0.011 (0.0736)	-0.139, 0.152	0.152	0.394	0.107 (0.0747)	-0.036, 0.254	1.435	0.142
Apple	base				base				base			
size 3-5	0.326 (0.0133)	0.3, 0.352 ^	24.474	0.000 **	0.246 (0.0127)	0.22, 0.27 ^	19.362	0.000 **	0.221 (0.0128)	0.196, 0.246 ^	17.281	0.000 **
Size 6+	0.589 (0.0291)	0.531, 0.646 ^	20.223	0.000 **	0.501 (0.027)	0.448, 0.554 ^	18.570	0.000 **	0.461 (0.0277)	0.406, 0.514 ^	16.635	0.000 **
Size 1s	base				base				base			
R-Squared (adj)	20.764%				55.185%				54.628%			
Mean Deviance	36,118.0				24,820.0				25,091.0			
Penalty	21.0				845.7				841.4			
DIC	36,139.0				25,666.0				25,933.0			
MAPE	6.237%				4.422%				4.465%			
Variance (between purchases)	0.318				0.187				0.190			
Variance (between housholds)					0.144				0.145			
between household t-stat (sig)					19.118(0)				19.459(0)			
Variance Partition Coeficient					43.409%				43.334%			

* significant 5%
** significant 1%
^ 95% Bayesian estimates do not include zero

Table 23: Model diagnostics and inference - fruit juice

A graphical representation of the parameters is shown in Fig 72. The diagnostics of the two hierarchical models are similar. There are some differences in the coefficients of some of the parameters especially relating to the Informational reinforcement within utilitarian group 2. The convergence charts are located in Figs 4-6 in the appendix and show the parameters have converged given the intertwined and stationary nature of the two chains. Also, the Gelman statistics in Table 24 confirm this convergence of the parameters.

	Hierarchical Vague		Hierarchical Informative		Non Hierarchical	
	Point Estimate	Upper CI	Point Estimate	Upper CI	Point Estimate	Upper CI
Constant	1	1	1	1	1	1
Log Price	1	1	1	1	1	1
Informational x Utilitarian Gp1	1	1	1	1	1	1
Informational x Utilitarian Gp2	1	1	1	1	1	1
SuperOwn x Informational	1	1	1	1	1	1
SuperOwn x Informational GP2	1	1	1	1	1	1
Chrsitmas	1	1	1	1	1	1
Chrsitmas Ut Gp2	1	1	1	1	1	1
Other fruit	1	1	1	1	1	1
Breakfast	1	1	1	1	1	1
Grape	1	1.01	1	1	1	1
Grapefruit	1	1	1	1	1	1
Mixed	1	1	1	1	1	1
Orange	1	1.01	1	1	1	1
Pineapple	1	1.01	1	1	1	1
Tomato	1	1.01	1	1	1	1
Vegetable	1	1	1	1	1	1
Vitamin	1	1	1	1	1	1
size 3-5	1	1	1	1	1	1
Size 6+	1	1	1	1	1	1

Table 24: Gelman convergence measures

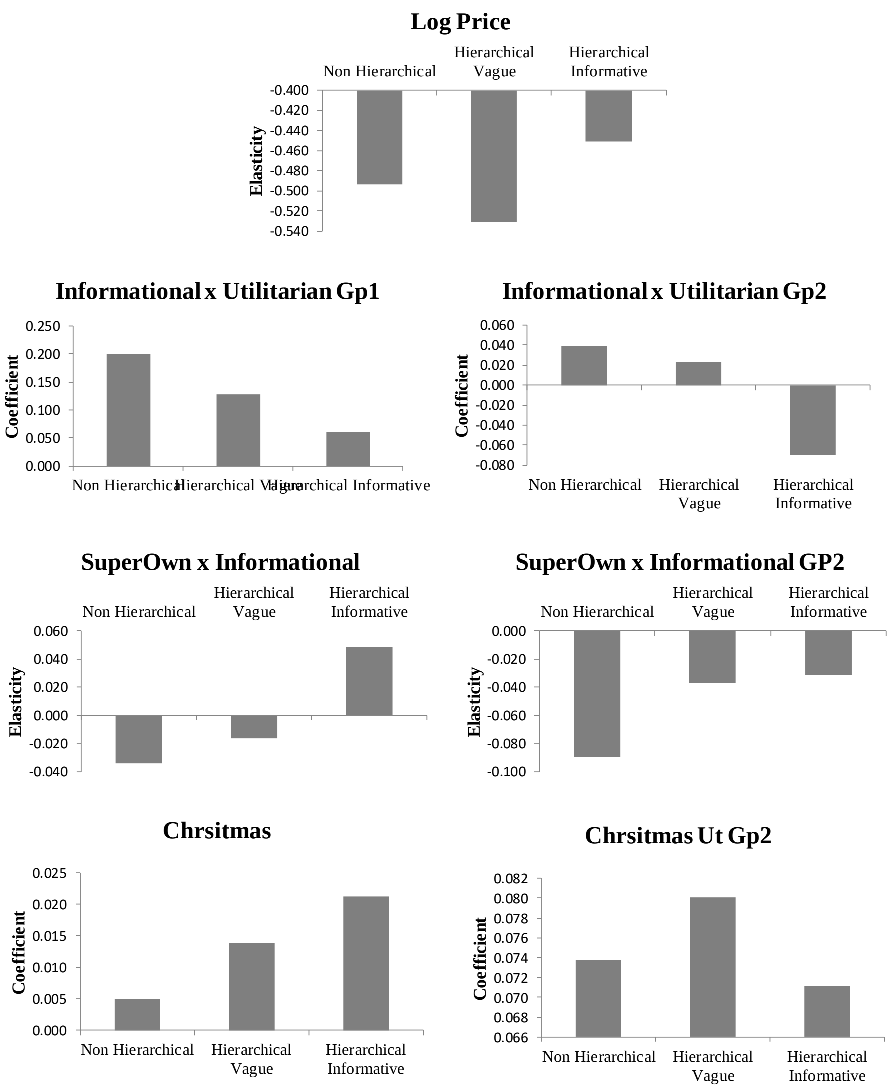

Figure 72: Parameter column charts

6.4.2 Coefficient discussion

6.4.2.1 Price Elasticity

Fig 73 shows the posterior density plot of the coefficient for the hierarchical (vague, and informative) and non-hierarchical models. Despite the difference being small in magnitude the box plots suggest this is statistically significant which demonstrates the difference in recognising the hierarchical structure of the data.

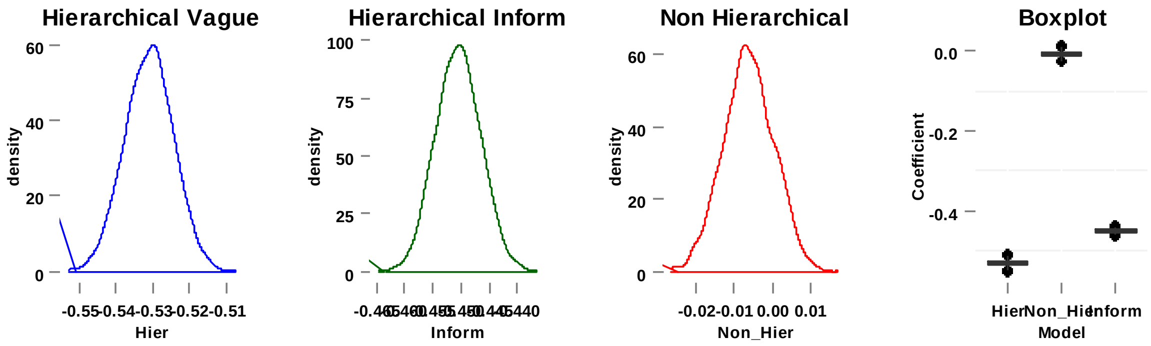

Figure 73: Price coefficients – fruit juice

From Fig 73 the point estimates of the price elasticity for the hierarchical vague, hierarchical informative and non-hierarchical models are -0.531 and -0.451 and -0.493 respectively. Ignoring the hierarchical structure has some impact on the elasticity of demand though this is small. Each coefficient's 95% Bayesian confidence interval of the posterior distribution does not include 0, the intervals being (-0.549, -0.513), (-0,462, -0.439) and (-0.510, -0.476) in turn. The frequentist t-statistics are -57.097, -79.035 and -57.337 respectively, all significant at $p<0.001$, hence little evidence to suggest the parameter is zero from a Bayesian or frequentist stance.

The magnitude of the coefficients is similar to past (non-hierarchical) studies (Chang, 2007; Oliveira-Castro *et al.*, 2006) and similar level of magnitude to the biscuit category. The nature of the hierarchy has resulted in a slightly different value of the parameter and although this is statistically significant due to the power of the test, it is unlikely to make a difference from a practical managerial perspective.

6.4.2.2 Informational reinforcement in the lower Utilitarian reinforcement group

As with the biscuit category, the informational variable is modelled within the lower utilitarian reinforcement group and an offset constructed to represent the higher informational group as this allows the statistical testing of the difference between the two utilitarian groups.

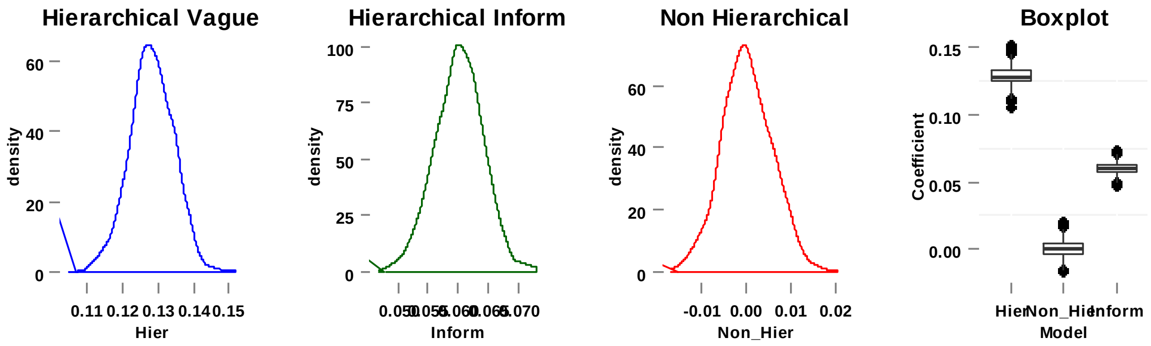

Figure 74: Informational reinforcement (lower Utilitarian group) - fruit juice

The posterior density and box plots of the informational reinforcement variable of the hierarchical and non-hierarchical models are shown in Fig 74. The point estimates of the models for the informational reinforcement variable within the lower utilitarian reinforcement group are 0.128, 0.060 and 0.200 respectively for the three models in the usual order. The posterior 95% Bayesian confidence intervals for the models in turn are (0.111, 0.145), (0.049, 0.072) and (0.183, 0.217), none of which include 0 suggesting the value is non-zero. The confidence intervals also do not overlap, again suggesting these estimations are significantly different for the estimates. The frequentist t-statistic for each respective model is 14.907, 10.379 and 24.118 respectively, all significant at $p<0.001$. Hence both models are suggesting a positive informational reinforcement is resulting in a positive effect on the category volume, above and beyond what can be explained by price.

6.4.2.3 Informational reinforcement in the lower Utilitarian reinforcement group (offset)

The informational reinforcement offset for the higher utilitarian group is discussed next and Fig 75 shows the density and boxplots for the posterior distribution of the parameters.

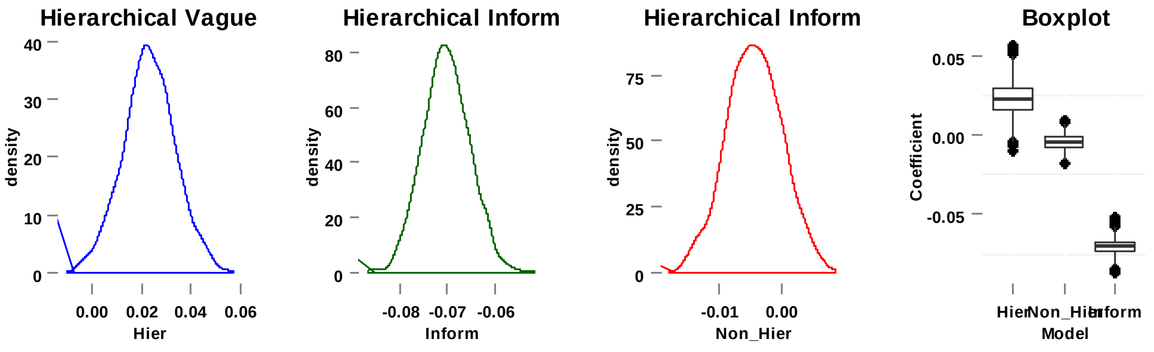

Figure 75: Informational reinforcement (higher Utilitarian group) - fruit juice

The point estimates for the parameter are 0.023 for the hierarchical vague, -0.070 for the hierarchical informative and 0.039 for the non-hierarchical models. The Bayesian confidence intervals are, in turn, (-0.006, 0.052), (-0,084, -0,057) and (0.009, 0.068) and the frequentist t-statistics for each are 1.54, -10.174 and 2.613 respectively. Hence there are different conclusions depending on the structure and prior distributions of each model, with the hierarchical vague suggesting the parameter is redundant (given the Bayesian confidence interval contains zero and the t-statistic is non-significant at p=0.1), the hierarchical informative model suggesting the parameter is negative and the non-hierarchical model suggesting the parameter is positive. This does underline the statement by Efron (2005) that Bayesian models can return differing results and why Leamer (1992), Rossi and Allenby (2003), Gelman (2010) says that it is important to understand the prior assumptions underpinning models.

6.4.2.4 Supermarket own brand x Informational reinforcement in the higher Utilitarian reinforcement group (offset)

Fig 76 shows the posterior distribution density and box plots of the informational and supermarket own interaction variable for all models.

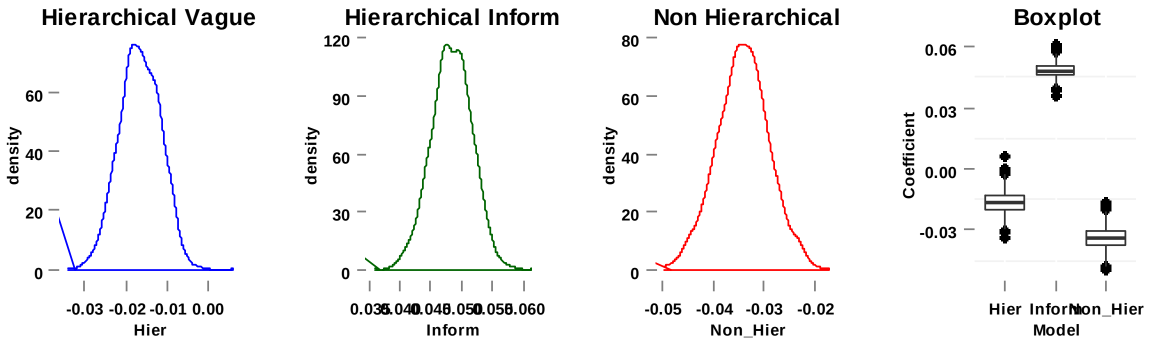

Figure 76: Supermarket own brand x Informational reinforcement (higher Utilitarian group) - fruit juice

The point estimates for the hierarchical and non-hierarchical models are -0.017, 0.048 and -0.034 respectively. The Bayesian confidence intervals of (-0.030, -0.003), (0.039, 0.057) and (-0.048, -0.020) none containing the value zero. There is disagreement between the models as to the sign of the coefficient with the hierarchical informative suggesting a positive impact on volume sales contrary to the other two models. The frequentist t-statistics of -2.371, 10.298 and -4.845, all p<0.03 reinforcing this disagreement. The disagreement is driven by the prior distribution of the hierarchical informative model which has a positive mean with a high precision. This, combined with the likelihood, is resulting in the positive estimate of the parameter for that specific model. This highlights the importance the prior distribution selection plays in model build.

6.4.2.5 Supermarket own brand x Informational reinforcement in the higher Utilitarian reinforcement group (offset)

Fig 77 shows the density and box plots of the posterior estimate of the coefficient.

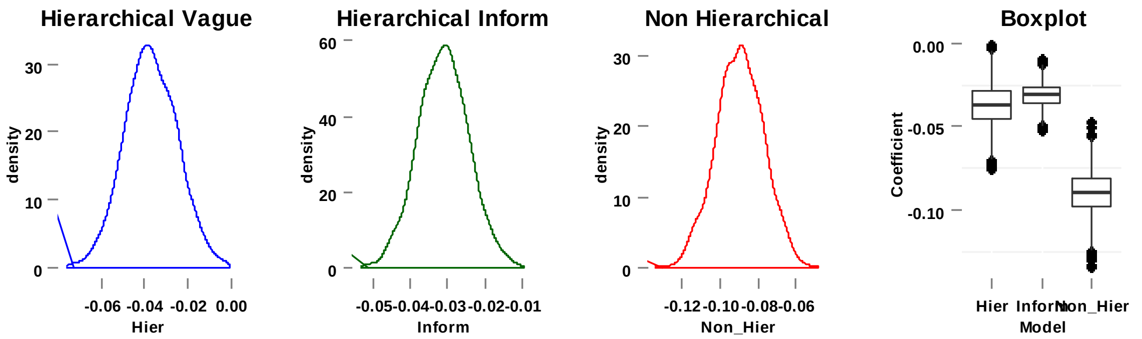

Figure 77: Supermarket own brand x Informational reinforcement (lower Utilitarian group) - fruit juice

The point estimates for the three models are, in turn, -0.037, -0.031 and -0.090. The posterior Bayesian confidence intervals of (-0.070, -0.003), (-0.050, -0.013) and (-0.124, -0.054) resulting in t-statistics of -2.17, -3.27 and -5.00 indicates these parameters are statistically significant to the model. This would suggest the effect is significantly higher within the informational criteria of the higher utilitarian reinforcement group when it comes to supermarket own brands. This means the volume of purchases will be lower for supermarket own brands which are seen as a higher utilitarian reinforcement group.

6.4.2.6 Christmas effect x Informational reinforcement in the lower Utilitarian reinforcement group

The density and box plots of the three models are shown in Fig 78 and their point estimates are 0.014, 0.021 and 0.005 respectively.

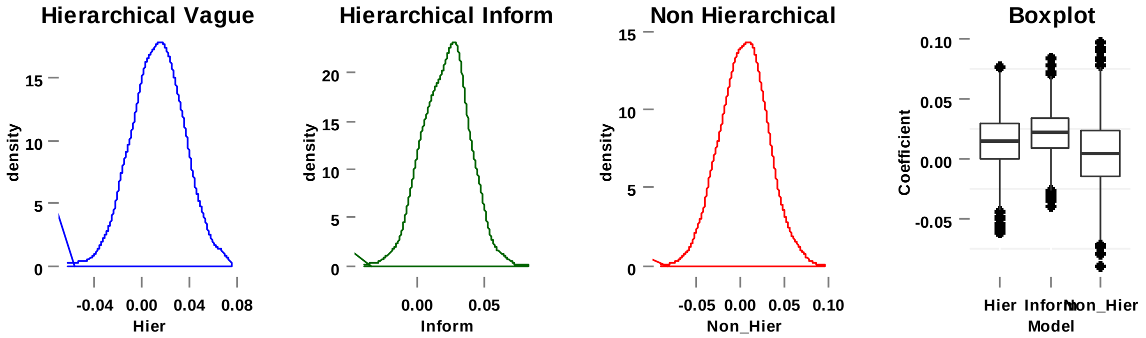

Figure 78: Christmas effect x Informational reinforcement (lower Utilitarian group) - fruit juice

The 95% Bayesian confidence intervals for the hierarchical vague, hierarchical informative and non-hierarchical models are (-0.045, 0.074), (-0.027 and 0.070) and (-0.073, 0.081) respectively, all of which contain the value zero suggesting the parameter is zero. The frequentist t-statistics of 0.454 (p=0.360) for the hierarchical vague, 0.858 (p=0.276) for the hierarchical informative and 0.125 (p=0.396) for the non-hierarchical also suggests there is no evidence the parameter is statistically significantly different from zero. Hence consumers' purchase volume of fruit juice associated with the lower Utilitarian group does not differ during the Christmas week, however there are fewer consumers purchasing which is the reason for the dip in volume in this period.

6.4.2.7 Christmas effect x Informational reinforcement in the higher Utilitarian reinforcement group (offset)

Fig 79 represents the density plots and box plots for the three models

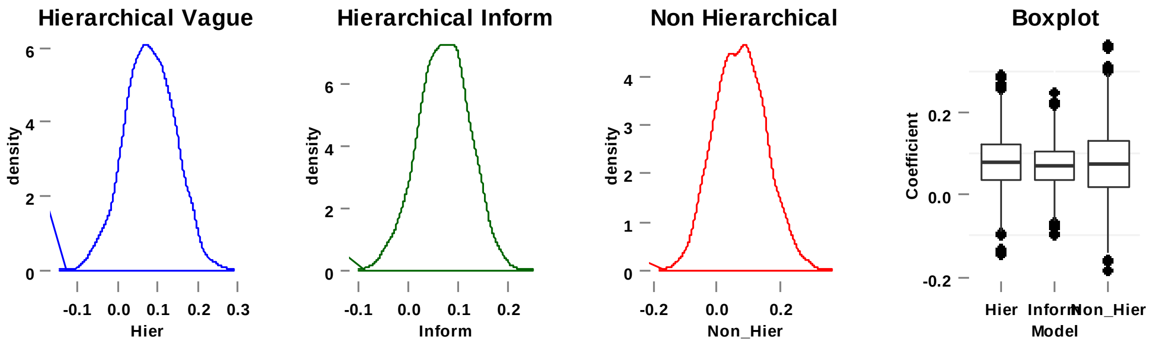

Figure 79: Christmas effect x Informational reinforcement (higher Utilitarian group) - fruit juice

The point estimates for the three models, in turn, are 0.080, 0.071 and 0.074 and hence similar values for the three models. However, in similar manner the Bayesian posterior distribution 95% confidence intervals of the three models straddle zero suggesting this parameter is not required within the model structure. The t-statistics are 0.454 (p=0.360), 0.858 (p=0.276) and 0.125 (p=0.396) all non-significant. Therefore, the volume purchased does not change for the higher utilitarian brands during the Christmas week.

6.4.2.8 Product Characteristic Variables

There is much agreement between the non-hierarchical and hierarchical models of the sign and significance of the product type variable, compared to *apple*, *other*, *grape* and *orange* are larger, *pineapple*, *grapefruit* and *tomato* smaller with little difference between *apple* and *other fruit*, *breakfast mixed*, *vegetable* and *vitamin* smaller.

The unit sales level of the number in pack is smaller than the increased pack sizes given their positive and significant coefficients, with the coefficient of 6+ being higher that 1-5. This is as expected given the coefficient is based on sales per transaction and likely larger pack sizes will account for more volume per purchase.

6.5 Yellow Fats

6.5.1 Model diagnostics

As with the previous categories, yellow fats analysis continues in the same manner. Figs 7-9 in the appendix shows the convergence charts for the posterior estimates of the focal variables. The overlapping chains and small bandwidth suggests the coefficients of both the hierarchical and non-hierarchical models have converged. The Gelman statistics in Table 25 confirm the convergence of the estimates. The small standard deviations of the estimates (relative to the parameter estimates) are small, again suggesting convergence.

	Hierarchical Vague		Hierarchical Informative		Non Hierarchical	
	Point Estimate	Upper CI	Point Estimate	Upper CI	Point Estimate	Upper CI
Constant	1	1	1	1	1	1
Log Price	1	1	1	1	1	1
Informational x Utilitarian Gp1	1	1	1	1	1	1
Informational x Utilitarian Gp2	1	1.01	1	1	1	1
SuperOwn x Informational	1	1	1	1	1	1
SuperOwn x Informational GP2	1	1	1	1	1	1
Christmas	1	1	1	1	1	1
Christmas x Ut Gp 2	1	1	1	1	1	1
Butter	1	1	1	1	1	1
Margarine	1	1	1	1	1	1
Low Reduced	1	1	1	1	1	1
Size 2+	1	1	1	1	1	1

Table 25: Gelman convergence measures – yellow fats

From the diagnostic variables shown in Table 26, the Bayesian mean deviance for the hierarchical vague model is 23,828, for the hierarchical informative model 24.242 and the non-hierarchical model 37,915. The respective penalties for the three models are 1,244, 1,239 and 13, resulting in a DIC of 25,072, 25,481 and 37,915. Therefore, despite the larger penalty incurred by the hierarchical structured models, the lower DIC suggests the models would better predict a replicated data set (Spiegelhalter *et al.*, 2002). From a frequentist stance, the R-squared (adjusted) values for the respective models are 58.119%, 57.529% and 30.967%, which implies the hierarchical structured models are accounting for a larger proportion of variance, even taking into account the penalty for the larger number of degrees of freedom required for these models. There is little difference between the two hierarchical models though the vague model does have the higher R-squared (adjusted) value. The MAPE for each model in turn is 4.251%, 4.301% and 6.050% again favouring the hierarchical structure given the lower mean average percentage error. The total residual variance for the three models is 0.127, 0.129 and 0.201 respectively, meaning the hierarchical structure has a lower residual variance than the non-hierarchical models. The hierarchical models' between household error variance values are 0.127 and 0.129 for the vague and informative models respectively. The t-statistics for the hierarchical variance coefficients are 23.714 and 23.386 (both significant at p<0.001) which result in variance partition coefficients of 37.428% and 37.476% respectively.

The range of diagnostics, both Bayesian and frequentist, suggest the hierarchical structure is contributing to the statistical representation of the data above and beyond the non-hierarchical structure. Despite this, the diagnostics of the non-hierarchical structure suggest this is also a good representation of the data and the study proceeds to discuss the coefficients associated with each model.

	Non Hierarchical					Hierarchical Vague					Hierarchical Informative				
	Beta	(SE posterior)	Bayes CI	t	sig	Beta	(SE posterior)	Bayes CI	t	sig	Beta	(SE posterior)	Bayes CI	t	sig
Constant	7.530	(0.02)	7.492, 7.569 ^	376.52	0.000 **	7.535	(0.0222)	7.492, 7.579 ^	339.42	0.000 **	7.625	(0.0153)	7.595, 7.655 ^	498.37	0.000 **
Log Price	-0.456	(0.0067)	-0.469, -0.443 ^	-68.04	0.000 **	-0.448	(0.007)	-0.462, -0.435 ^	-64.04	0.000 **	-0.447	(0.0041)	-0.455, -0.439 ^	-109.12	0.000 **
Informational x Utilitarian Gp1	0.173	(0.005)	0.164, 0.183 ^	34.62	0.000 **	0.157	(0.0053)	0.146, 0.168 ^	29.62	0.000 **	0.106	(0.0034)	0.099, 0.112 ^	31.06	0.000 **
Informational x Utilitarian Gp2	-0.018	(0.0041)	-0.026, -0.01 ^	-4.39	0.000 **	-0.0007	(0.0042)	-0.009, 0.007	-0.17	0.393	-0.047	(0.0028)	-0.052, -0.041 ^	-16.64	0.000 **
SuperOwn x Informational	-0.057	(0.0062)	-0.069, -0.045 ^	-9.21	0.000 **	-0.033	(0.0061)	-0.045, -0.021 ^	-5.44	0.000 **	-0.044	(0.0041)	-0.052, -0.035 ^	-10.68	0.000 **
SuperOwn x Informational GP2	0.052	(0.0107)	0.031, 0.073 ^	4.84	0.000 **	0.004	(0.0104)	-0.017, 0.025	0.39	0.369	-0.013	(0.0078)	-0.029, 0.002	-1.71	0.093
Christmas	-0.063	(0.0267)	-0.114, -0.011 ^	-2.37	0.024 *	-0.044	(0.0215)	-0.085, -0.002 ^	-2.04	0.050 *	-0.051	(0.017)	-0.083, -0.016 ^	-2.98	0.005 **
Christmas x Ut Gp 2	0.072	(0.0626)	-0.05, 0.196	1.16	0.204	0.030	(0.0509)	-0.069, 0.13	0.59	0.335	-0.013	(0.0401)	-0.094, 0.065	-0.31	0.380
Butter	-0.308	(0.0075)	-0.323, -0.294 ^	-41.09	0.000 **	-0.339	(0.008)	-0.355, -0.323 ^	-42.36	0.000 **	-0.305	(0.0073)	-0.319, -0.29 ^	-41.73	0.000 **
Margarine	-0.187	(0.0074)	-0.201, -0.173 ^	-25.24	0.000 **	-0.188	(0.0079)	-0.204, -0.173 ^	-23.85	0.000 **	-0.194	(0.0077)	-0.209, -0.179 ^	-25.21	0.000 **
Low Reduced	-0.122	(0.0084)	-0.138, -0.105 ^	-14.49	0.000 **	-0.113	(0.0089)	-0.13, -0.095 ^	-12.65	0.000 **	-0.096	(0.009)	-0.113, -0.078 ^	-10.69	0.000 **
Blended spreads	base					base					base				
Size 2+	0.427	(0.0386)	0.353, 0.503 ^	11.05	0.000 **	0.429	(0.0342)	0.361, 0.497 ^	12.54	0.000 **	0.390	(0.034)	0.323, 0.455 ^	11.46	0.000 **
Size 1s	base					base					base				
R-Squared (adj)	30.967%					58.119%					57.529%				
Mean Deviance	37,902.0					23,828.0					24,242.0				
Penalty	13.1					1,244.0					1,239.0				
DIC	37,915.0					25,072.0					25,481.0				
MAPE	6.050%					4.251%					4.301%				
Variance (between purchases)	0.201					0.127					0.129				
Variance (between housholds)						0.076					0.077				
between household t-stat (sig)						23.714(0)					23.386(0)				
Variance Partition Coeficient						37.428%					37.476%				

* significant 5%
** significant 1%
^ 95% Bayesian estimates do not include zero

Table 26: Model diagnostics and inference - yellow fats

Fig 80 represents a graphical view of the coefficient point estimates of the posterior distributions of the focal variables of the Yellow Fats category. From a visual perspective, it seems the offset variables have the most conflicting views of the parameter estimates given the pattern of the column charts.

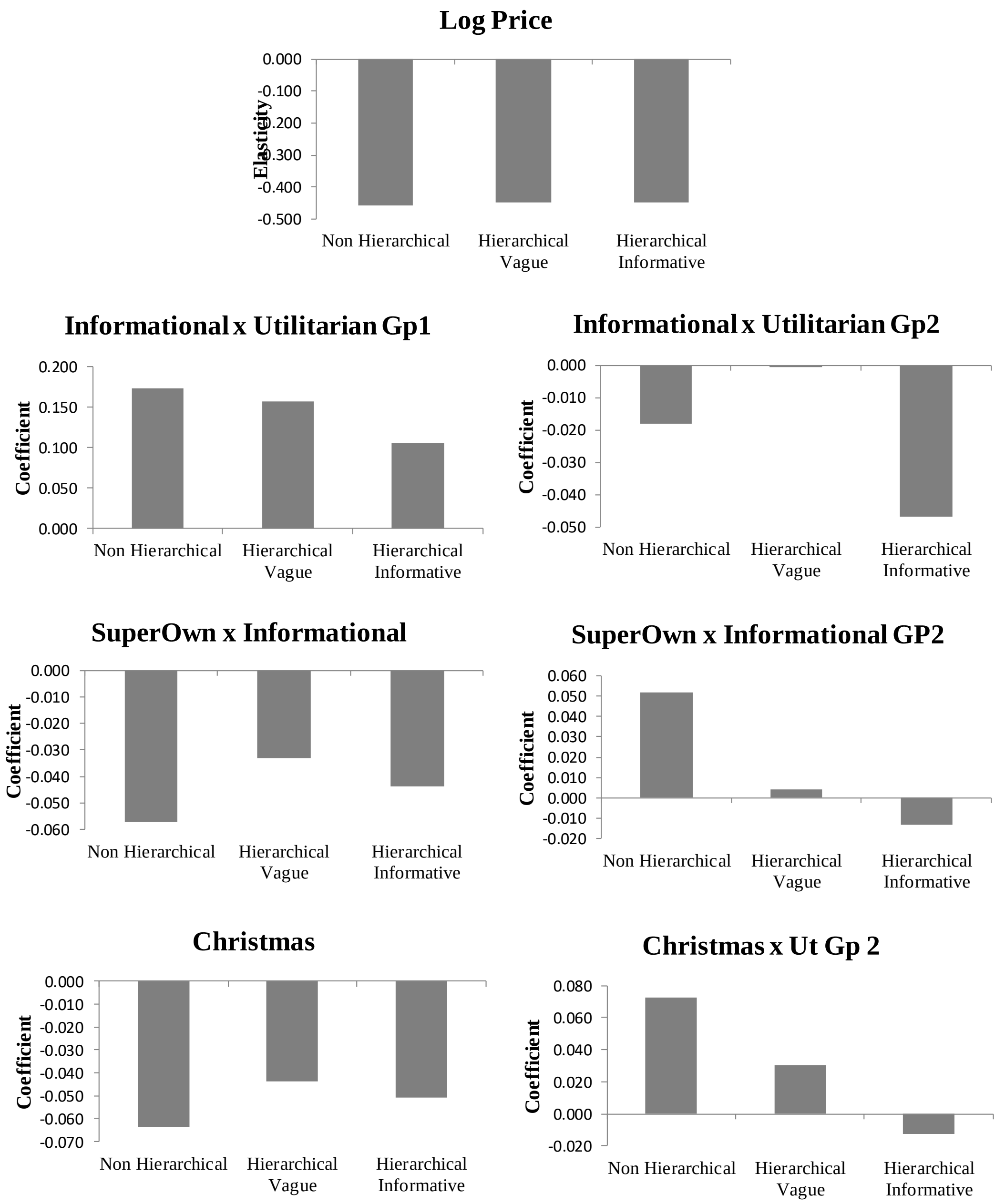

Figure 80: Parameter column charts – yellow fats

In order to determine whether these are statistically different requires a more detailed analysis of the parameters' posterior distribution estimates.

6.5.2 Coefficient discussion

6.5.2.1 Price Elasticity

The point estimate of the elasticity coefficients' posterior distribution for the hierarchical vague, hierarchical informative and non-hierarchical models are -0.448, -0.447 and -0.456 in turn, which are of similar magnitude to each other, to other categories within this study and to results from other studies[2] (Chang, 2007; Oliveira-Castro *et al.*, 2006).

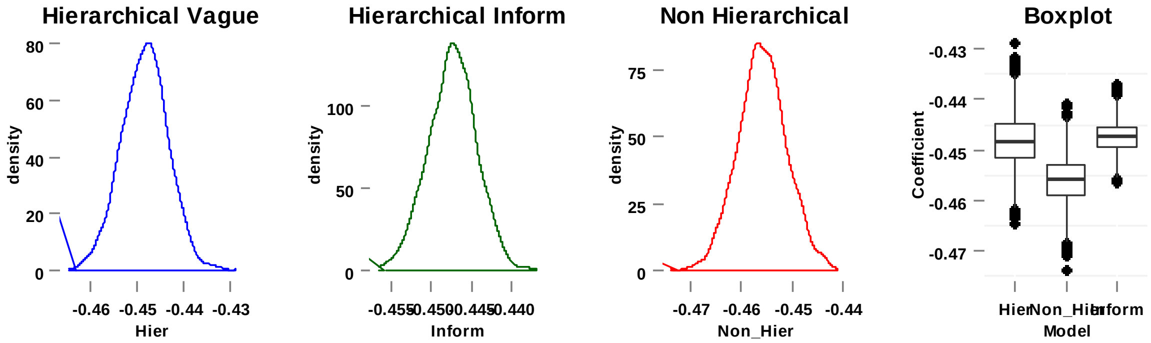

Figure 81: Price coefficients - yellow fats

Fig 81 shows the Bayesian posterior distribution confidence intervals for the hierarchical and non-hierarchical models are (-0.461, -0.435), (-0.455, -0.439) and (-0.469, -0.443) in the usual order, hence no inclusion of the 0 value in any interval suggesting the parameter is non-zero. This view is strengthened by the large frequentist t-statistics of (-64.4, -109.1 and -60.0 in respective order, all $p<0.001$). Despite a lack of overlap in the confidence intervals of the hierarchical and non-hierarchical models, the estimates are very similar on a practical level and the large sample size is contributing to the tight estimates in the probability of the Bayesian confidence intervals.

6.5.2.2 Informational reinforcement in the lower Utilitarian reinforcement group

The posterior distributions of the informational reinforcement are shown in the density and boxplots in Fig 82. The point estimates for the hierarchical and non-hierarchical models in the usual order are 0.157, 0.106 and 0.173.

[2] Based on non-hierarchical studies

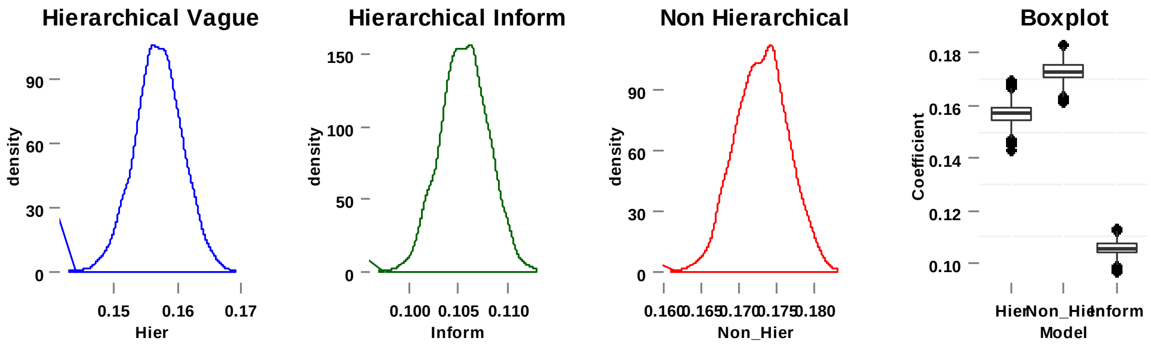

Figure 82: Informational reinforcement (lower Utilitarian group) - yellow fats

The Bayesian posterior distribution confidence intervals are (0.146, 0.168), (0.099, 0.112) and (0.164, 0.183) respectively, none of which containing the value 0, suggesting the parameter is statistically required. The respective frequentist t-statistics for the parameter are 29.6, 31.0 and 34.6 respectively, each with $p<0.001$ reinforcing the statistical requirement of the parameter. Therefore, the informational reinforcement is positively influencing volume per purchase above and beyond what can be accounted for by price. However, the nature of the confidence intervals suggests the ignorance of the hierarchical structure means this reinforcement is smaller than when taking the hierarchy of the data into account.

6.5.2.3 Informational reinforcement in the higher Utilitarian reinforcement group (offset)

Fig 83 shows the posterior density and box plots for the informational reinforcement variable within the higher utilitarian reinforcement group, hence is an offset to the informational reinforcement within the lower utilitarian group.

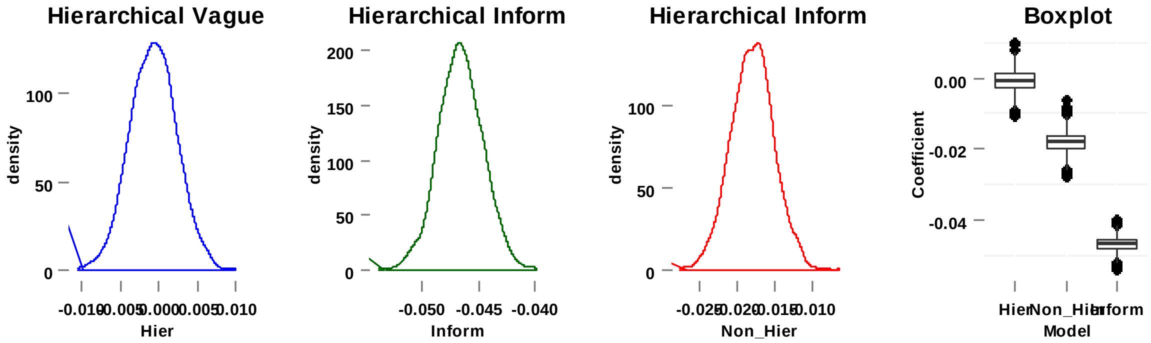

Figure 83: Informational reinforcement (higher Utilitarian group) - yellow fats

The point estimates of the Bayesian posterior coefficients are -0007, -0.047 and -0.018 in the usual order for the hierarchical (vague and informative) and non-hierarchical models respectively. The respective Bayesian posterior confidence intervals are (-0.009, 0.007), (-0.052, -0.041) and (-0.026, -0.01). The frequentist t-statistic in turn are -0.17 ($p=0.393$), -16.6 ($p<0.001$) and -4.39 ($p<0.001$). The Bayesian confidence interval of the hierarchical vague model contains zero and hence with 95% probability it cannot be concluded this parameter is non-zero. The frequentist t-statistic brings the same conclusion. However, for the hierarchical informative and non-hierarchical models, both Bayesian and frequentist measures suggest the value of this offset variable is negative, hence suggesting the informational reinforcement in the higher utilitarian reinforcement group is lower than in the lower utilitarian group. The negative mean and large precision of the prior distribution of the hierarchical informative model is influencing the parameter to be negative. The hierarchical vague model does not have this strong precision and is influenced more by the likelihood derived from the data, suggesting the parameter is zero. This is a further example of the implication of model functional form and prior distribution selection has on the posterior estimates of the model parameters.

Therefore, the informational reinforcement variable within the BPM is contributing to the volume per purchase of the yellow fats category above and beyond what can be accounted for by price alone. Whether there is a significant difference in how volume is influenced by this informational reinforcement between the upper and lower utilitarian reinforcement groups would depend on which model structure and which prior distribution is preferred. The difference associated with the prior is an example of the fundamental disagreement which has existed historically between the Bayesian and frequentist arguments. However, as O'Hagan (1994) and Duncan *et al.,* (1996) would argue, having these informed discussions at the beginning of a model build where the level of uncertainty around a parameter can be included mathematically into a model is more useful than making decisions post hoc as to the validity of the parameter.

6.5.2.4 Supermarket own brand x Informational reinforcement in the lower Utilitarian reinforcement group

As with the previous set of variables, this supermarket own effect on informational reinforcement is categorised within the lower Utilitarian reinforcement group as a base measure and the upper Utilitarian reinforcement group as an offset. This allows the statistical consideration of the difference between them.

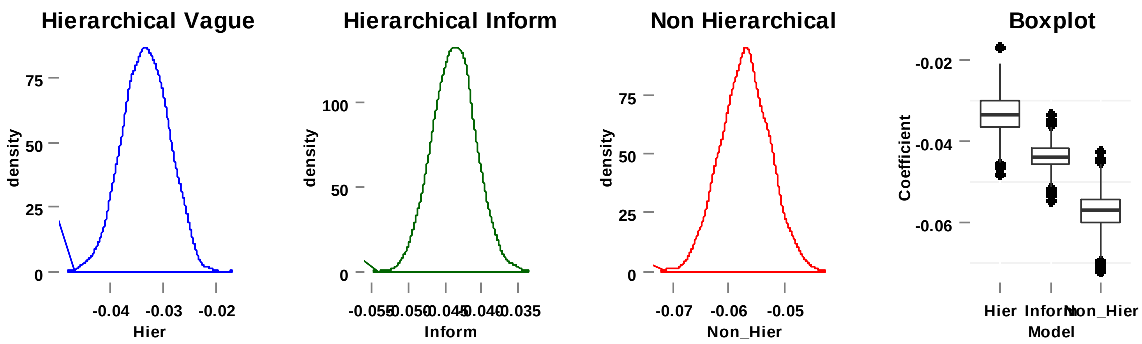

Figure 84: Supermarket own brand x Informational reinforcement (lower Utilitarian group) - yellow fats

Fig 84 shows the density and box plots for the Bayesian posterior estimates. The point estimates in turn are -0.033, -0.044 and -0.057 for the hierarchical and non-hierarchical models in the usual order. The Bayesian posterior confidence intervals for the three respective models are (-0.045, -0.021), (-0.052, -0.035) and (-0.069, -0.045) all of which do not contain zero. Also, the significant t-statistics (-5.44, -10.68, -9.21, all $p<0.001$) suggest the parameter is non-zero for all three models. This suggests a negative effect on volume for this parameter. Considering the hierarchical confidence intervals, it is noted the vague model's confidence interval lies within the confidence interval of the informative model. This shows some agreement between the informative prior distribution and the likelihood from the data.

All three models suggest a supermarket own brand's informative reinforcement has a negative effect on volume of purchase, within the lower utilitarian reinforcement group. This could mean consumers shopping for a lower equity product in a low utilitarian reinforcement requirement may be put off if the product has a higher informational reinforcement associated within it. This could be due to the conflicting needs of value versus informative reinforcement nature of the product. Hence if a product is aimed at a supermarket own brand and is targeted to the lower utilitarian group then it is actually beneficial to associate lower informational reinforcement scores to the product, if volume maximisation is the goal.

6.5.2.5 Supermarket own brand x Informational reinforcement in the higher Utilitarian reinforcement group (offset)

Fig 85 shows the density and box plots for the offset of the Informational Reinforcement variable within the higher Utilitarian reinforcement group with regards to supermarket own brands. The offset is versus the same variable but in the lower Utilitarian reinforcement group.

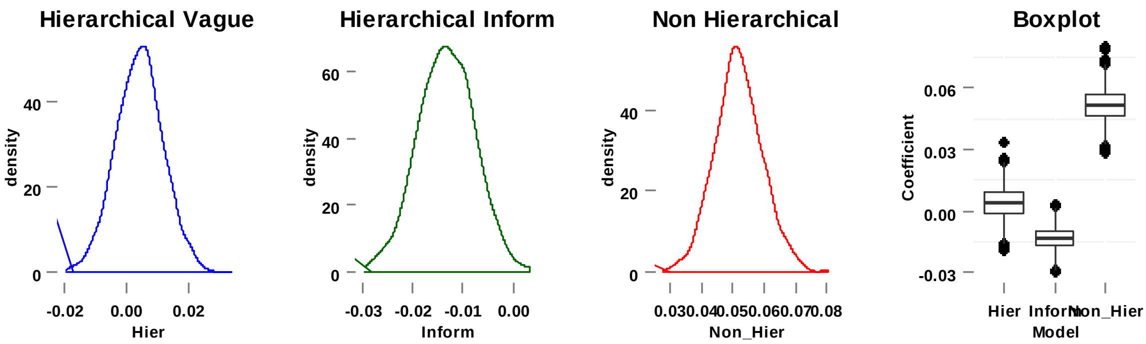

Figure 85: Supermarket own brand x Informational reinforcement (higher Utilitarian group) - yellow fats

The Bayesian point estimates for the three models in turn are 0.004, -0.013 and 0.052 in the usual respective order. The Bayesian posterior confidence intervals for both the hierarchical variants of the model straddle zero, i.e. (-0.017, 0.025) and (-0.029, 0.002). Also, their t-statistics show no evidence to reject the parameter being zero ($t=0.39$, $p=0.369$; $t= -1.81$, $p=0.09$ respectively) which suggest the offset of the variable extended to the higher Utilitarian group is not significantly different from the lower Utilitarian group. However, when considering the non-hierarchical model, the Bayesian confidence intervals are (0.031, 0.073) and frequentist t-statistic of 4.84 ($p<0.001$) which suggests the higher Utilitarian group extension of the variable is statistically significantly higher than the lower group. This demonstrates the difference in results observed if the structure of the data is not considered.

6.5.2.6 Christmas effect x Informational reinforcement in the lower Utilitarian reinforcement group

The density plots and box plots of the hierarchical and non-hierarchical coefficients of the Christmas holiday week are shown in Fig 86 and their point estimates are -0.044, -0.051 and -0.063 respectively.

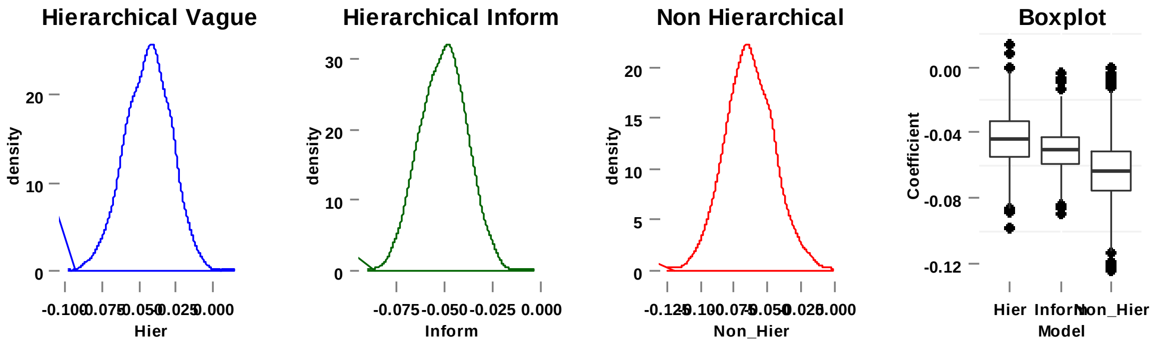

Figure 86: Christmas effect x Informational reinforcement (lower Utilitarian group) - yellow fats

The Bayesian confidence intervals are (-0.085, -0.002), (-0.083, -0.016) and (-0.114, -0.011) for the models in turn which would illustrate the volume per purchase is negatively impacted within this period. The interpretation of the frequentist t-statistics would agree with t= -2.04 (p=0.05), p= -2.98 (p=0.005) and t= -2.37 (p=0.024). This would suggest for the lower Utilitarian reinforcement group within the yellow fat category volume is significantly lower per purchase within the defined Christmas week than average purchase rates at other times of the year.

6.5.2.7 Christmas effect x Informational reinforcement in the higher Utilitarian reinforcement group (offset)

The Bayesian posterior point estimates for the coefficient of the three models are shown in the density and box plots in Fig 87.

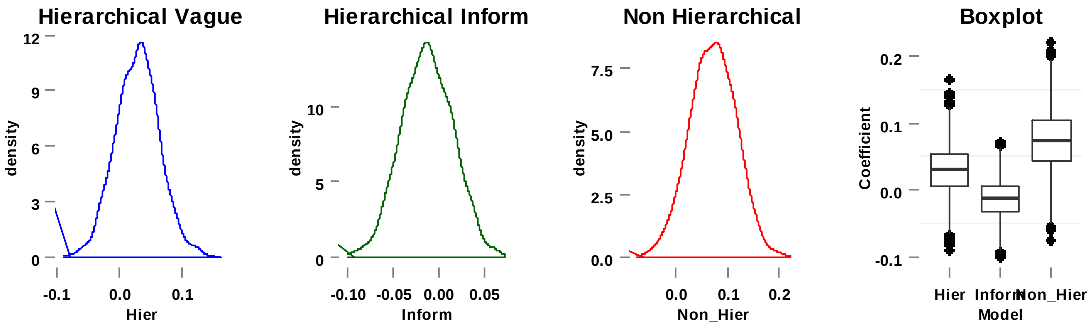

Figure 87: Christmas effect x Informational reinforcement (higher Utilitarian group) - yellow fats

Each Bayesian confidence interval straddles zero, (-0.069, 0.13), (-0.094, 0.065) and (-0.05, 0.196) respectively and all have small t-statistics which are not statistically significant ($t=0.59$, $p=0.34$; $t= -0.31$, $p=0.38$ and $t=1.16$, $p=0.20$ respectively) for the hierarchical vague, hierarchical informative and non-hierarchical models. This implies that the Christmas week offset for the higher Utilitarian reinforcement group is no different from that of the lower Utilitarian group, hence volume level per purchase is not impacted during the Christmas week within the higher Utilitarian group above and beyond what is observed in the lower Utilitarian group.

In summary, the results would lead us to conclude that volume per purchase within the yellow fats category is negatively impacted during the Christmas week, however this does not differentiate as to whether the product is of a lower or higher Utilitarian group.

6.5.2.8 Characteristic Variables

There is considerable agreement across both the non-hierarchical and hierarchical models in the direction and significance of the characteristic variables and all are significant under the Bayesian and frequentist inference statistics. Compared to *Blended spreads*, all other variants have a negative coefficient which is statistically significant across all models and for both Bayesian and frequentist inference

As seen in other categories, the unit items in pack have a smaller volume per transaction than the larger sizes. Here is no exception, with the 2+ category yielding a positive and statistically significant coefficient across all models and for both Bayesian and frequentist inferences.

6.6 Baked Beans

6.6.1 Model Diagnostics

Figs 10-12 in the appendix shows the convergence plots for the two hierarchical and the non-hierarchical models suggesting the parameters have converged in all cases. The Gelman statistics for both halves of the chains are close to 1 showing convergence, see Table 27.

	Hierarchical Vague		Hierarchical Informative		Non Hierarchical	
	Point Estimate	Upper CI	Point Estimate	Upper CI	Point Estimate	Upper CI
Constant	1	1	1	1	1	1
Log Price	1	1	1	1	1	1
Informational x Utilitarian Gp1	1	1	1	1	1	1
Informational x Utilitarian Gp2	1	1	1	1	1	1
SuperOwn x Informational	1	1	1	1	1	1
SuperOwn x Informational x UT Gp2	1	1	1	1	1	1
Christmas	1	1.01	1	1	1	1
Christmas UT Gp2	1	1	1	1	1	1
Beans Plus	1	1	1	1	1	1
Tomato	1	1	1	1	1	1
Healthy	1	1	1	1.01	1	1.01

Table 27: Gelman convergence measures - beans

The model diagnostics are displayed in Table 28 overleaf. The Bayesian inference measures show the Mean Deviance figures at 12,217, 12,643 and 19,464 respectively for the hierarchical vague, hierarchical informative and non-hierarchical models. The penalty measure for the three in turn is 763.6, 760.0 and 14.1, resulting in a DIC of 12,981, 13,404 and 19,478 respectively.

The R-squared (adjusted) values are 76.697%, 75.925 and 58.038% for the three models respectively and the MAPE statistics are 4.274%, 4.336% and 5.840% respectively. All these measures indicate the hierarchically structured models are performing statistically better than the non-hierarchical structure. There is a similarity between the two hierarchical models in terms of diagnostics though the vague hierarchical models have consistently better diagnostics than the informative model. The residual values for the three respective models are 0.143, 0.148 and 0.243 which confirms the hierarchical structure explaining a greater proportion on the variance of the data than the non-hierarchical structure. The hierarchical variance term for the vague and informative models is 0.105 and 0.109 respectively, which derive variance partition coefficients of 42.374% and 42.375%. The associated t-statistics are 18.67 and 18.13, (both significant at $p<0.001$) which indicate the between household variance term is statistically significant within the model structure. This concludes the hierarchical nature of the model is benefitting the model. However, diagnostically, there is little difference between the vague and informative models.

	Non Hierarchical				Hierarchical Vague				Hierarchical Informative			
	Beta (SE posterior)	Bayes CI	t	sig	Beta (SE posterior)	Bayes CI	t	sig	Beta (SE posterior)	Bayes CI	t	sig
Constant	7.542 (0.0192)	7.505, 7.579 ^	392.81	0.000 **	7.334 (0.0254)	7.284, 7.383 ^	288.72	0.000 **	7.363 (0.0198)	7.324, 7.403 ^	371.88	0.000 **
Log Price	-0.571 (0.0112)	-0.592, -0.549 ^	-50.96	0.000 **	-0.476 (0.0118)	-0.499, -0.453 ^	-40.35	0.000 **	-0.443 (0.0068)	-0.456, -0.43 ^	-65.15	0.000 **
Informational x Utilitarian Gp1	0.002 (0.0068)	-0.011, 0.015	0.29	0.382	0.029 (0.0076)	0.014, 0.044 ^	3.78	0.000 **	-0.026 (0.0049)	-0.036, -0.016 ^	-5.27	0.000 **
Informational x Utilitarian Gp2	0.041 (0.0065)	0.029, 0.054 ^	6.32	0.000 **	0.005 (0.0064)	-0.008, 0.017	0.80	0.290	-0.075 (0.0046)	-0.084, -0.066 ^	-16.26	0.000 **
SuperOwn x Informational	-0.113 (0.0081)	-0.129, -0.097 ^	-13.91	0.000 **	-0.088 (0.0096)	-0.107, -0.07 ^	-9.21	0.000 **	-0.012 (0.0065)	-0.025, 0	-1.89	0.067
SuperOwn x Informational x UT	-0.011 (0.0112)	-0.033, 0.011	-0.96	0.253	0.031 (0.0115)	0.009, 0.054 ^	2.71	0.010 *	0.008 (0.009)	-0.009, 0.026	0.92	0.261
Christmas	-0.030 (0.0456)	-0.118, 0.058	-0.65	0.322	0.010 (0.0359)	-0.063, 0.08	0.27	0.385	0.003 (0.0306)	-0.057, 0.063	0.08	0.397
Christmas UT Gp2	0.178 (0.094)	-0.004, 0.365	1.90	0.066	0.133 (0.0743)	-0.011, 0.277	1.78	0.081	0.123 (0.0628)	0.001, 0.251 ^	1.95	0.059
Beans Plus	0.005 (0.012)	-0.019, 0.028	0.38	0.372	0.007 (0.0113)	-0.016, 0.029	0.63	0.327	0.010 (0.0115)	-0.013, 0.033	0.84	0.280
Tomato	-0.007 (0.0111)	-0.028, 0.015	-0.59	0.336	0.012 (0.0103)	-0.008, 0.032	1.17	0.202	0.015 (0.0104)	-0.006, 0.036	1.46	0.137
Healthy	-0.015 (0.019)	-0.052, 0.022	-0.79	0.292	-0.008 (0.018)	-0.042, 0.027	-0.44	0.361	0.001 (0.0186)	-0.035, 0.037	0.06	0.398
Flavours	-0.051 (0.03)	-0.11, 0.009	-1.69	0.096	-0.022 (0.0266)	-0.075, 0.028	-0.83	0.282	-0.020 (0.0262)	-0.073, 0.032	-0.76	0.299
Beans Only	base				base				base			
Size 2+	1.132 (0.0114)	1.109, 1.154 ^	99.25	0.000 **	0.991 (0.0122)	0.967, 1.015 ^	81.25	0.000 **	0.978 (0.0121)	0.954, 1.001 ^	80.85	0.000 **
Size 1s	base				base				base			
R-Squared (adj)	58.038%				76.697%				75.925%			
Mean Deviance	19,464.0				12,217.0				12,643.0			
Penalty	14.1				763.6				760.0			
DIC	19,478.0				12,981.0				13,404.0			
MAPE	5.840%				4.274%				4.336%			
Variance (between purchases)	0.243				0.143				0.148			
Variance (between housholds)					0.105				0.109			
between household t-stat (sig)					18.67(0)				18.126(0)			
Variance Partition Coeficient					42.374%				42.375%			

* significant 5%
** significant 1%
^ 95% Bayesian estimates do not include zero

Table 28: Model diagnostics and inference - beans

The study continues with a discussion on the coefficients of the model which are shown numerically in Table 28 with the focal parameters shown graphically in Fig 88.

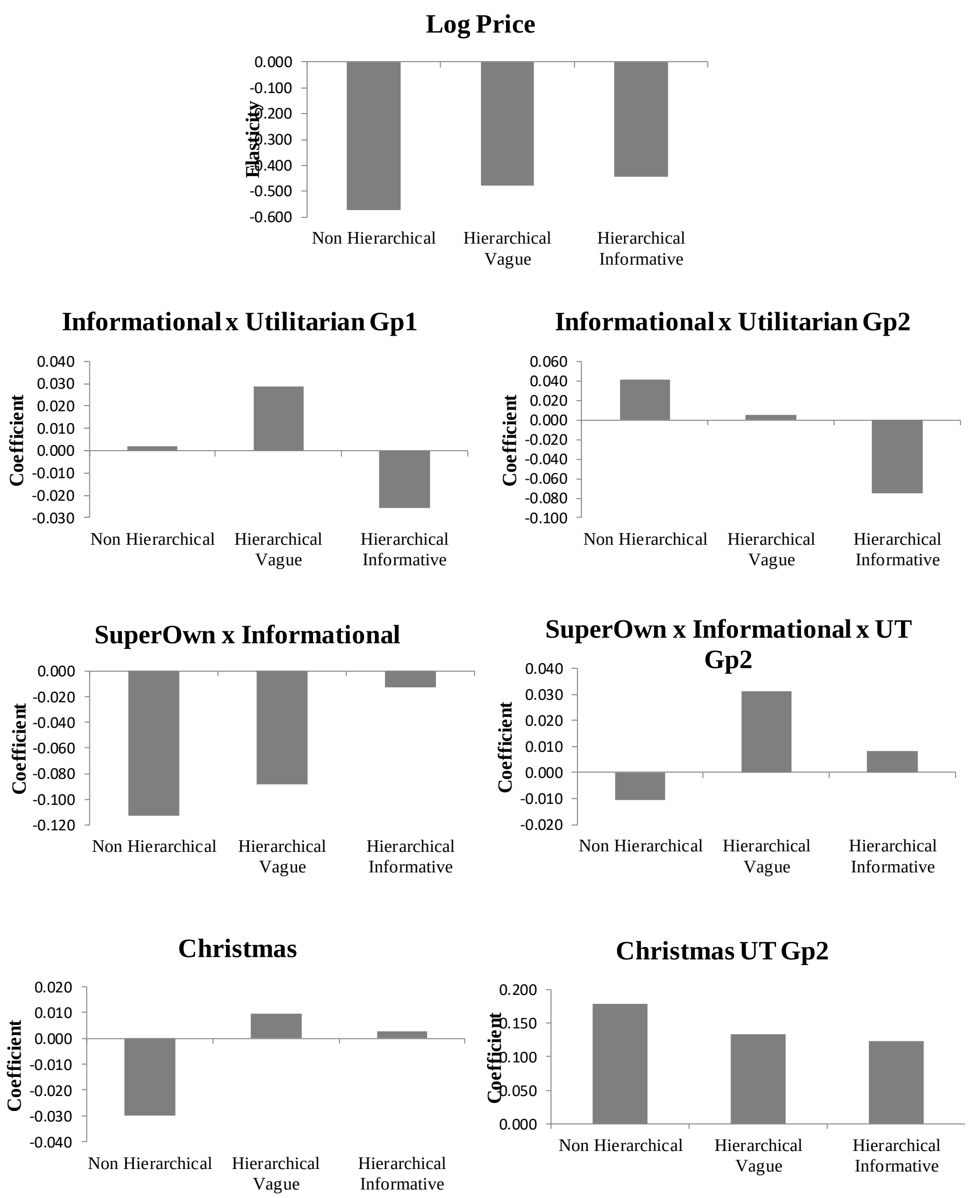

Figure 88: Parameter column charts - beans

6.6.2 Coefficient discussion

6.6.2.1 Price Elasticity

Fig 89 shows the posterior distribution of the price elasticity measure as density and box plots.

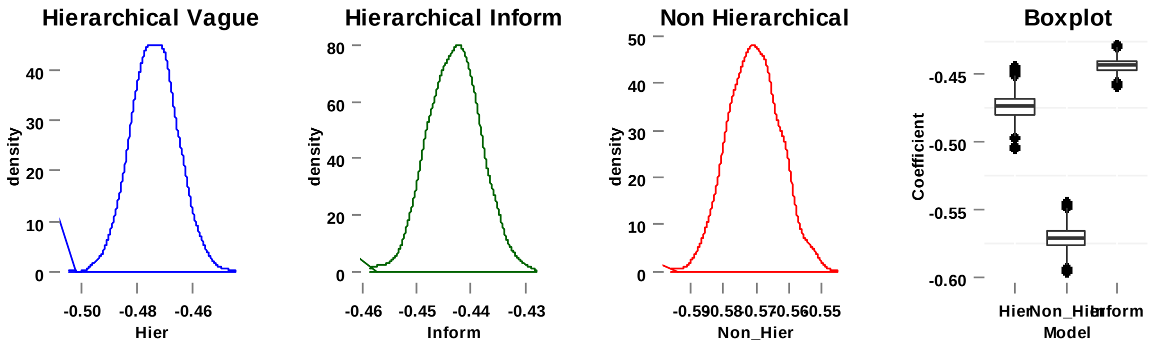

Figure 89: Price coefficients - beans

The price elasticity point estimate coefficient for the hierarchical vague, informative and non-hierarchical models is -0.476, -0.443 and -0.571 respectively which are similar in magnitude to each other, other categories and other studies[3]. The Bayesian confidence interval for the respective models are (-0.499, -0.453), (-0.456, -0.430) and (-0.592, -0.549) hence no inclusion of the value zero for any model. All frequentist t-statistics are large in magnitude (-40.35, -65.15 and -50.96), hence are statistically significant ($p<0.001$) giving strong evidence to reject the hypothesis the parameter is zero.

6.6.2.2 Informational reinforcement in the lower Utilitarian reinforcement group

Fig 90 shows the density and boxplots of the posterior distribution of the informational reinforcement variable for the lower utilitarian reinforcement group (the base group).

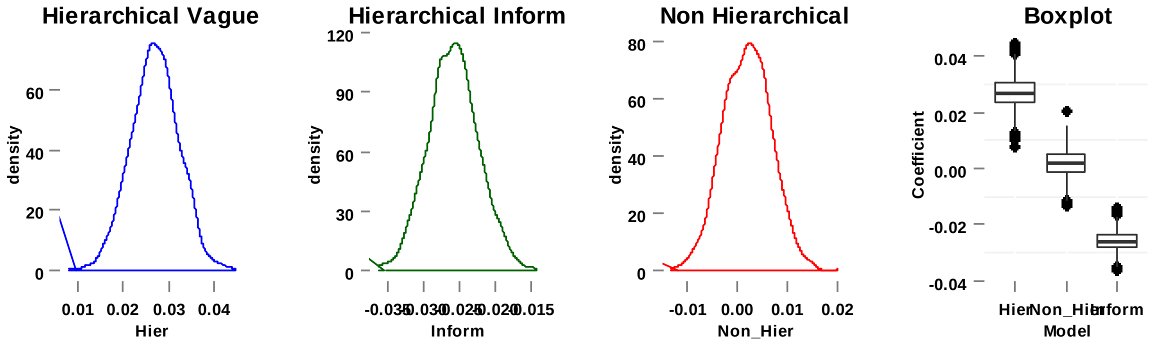

Figure 90: Informational reinforcement (lower Utilitarian group) - beans

[3] Non hierarchical studies

The point estimates for each model in the usual order are 0.029, -0.026 and 0.002 with respective Bayesian posterior confidence intervals of (0.014, 0.044), (-0.036, -0.016) and (-0.011, 0.015) and t-statistics of 3.78 ($p<0.001$), -5.265 ($p<0.001$) and 0.294 ($p=0.382$). Therefore, each model is deriving a different interpretation of the coefficient with the hierarchical vague model suggesting a statistically significant positive effect, the hierarchical informative implying a statistically significant negative effect and the non-hierarchical model implying the parameter is zero valued. The negative mean and large precision of the informative model is influencing the parameter for that model, this reinforces the need to understand the structure of the model being built and the prior knowledge which is built into the prior distribution of the model, since the results can be very different depending on these factors.

6.6.2.3 Informational reinforcement in the higher Utilitarian reinforcement group (offset)

Fig 91 shows the density and boxplots of the informational variable interaction with the higher utilitarian group.

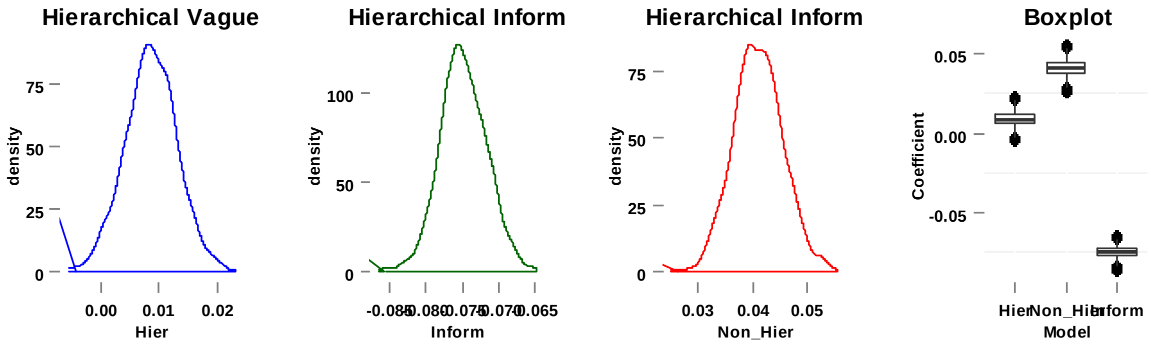

Figure 91: Informational reinforcement (higher Utilitarian group) - beans

The point estimates for the models in the usual order are 0.005, -0.075 and -0.011. When considering the Bayesian confidence intervals of each model, (-0.008, 0.017), (-0.084, -0.066) and (0.029, 0.054) and also their associated t-statistics ($t=0.80$, $p=0.29$; $t=-16.261$, $p<0.001$; $t=6.32$, $p<0.001$) there is again conflicting estimates from the three models. The hierarchical vague suggests this offset parameter is zero, the hierarchical informative suggests its value is negative and the non-hierarchical a positive relationship. The informative model is being influenced by a strong negative prior derived from the preliminary analysis.

6.6.2.4 Supermarket own brand x Informational reinforcement in the lower Utilitarian reinforcement group

Fig 92 shows the density and boxplots of the posterior estimate of the interaction of informational and supermarket own brand indicator.

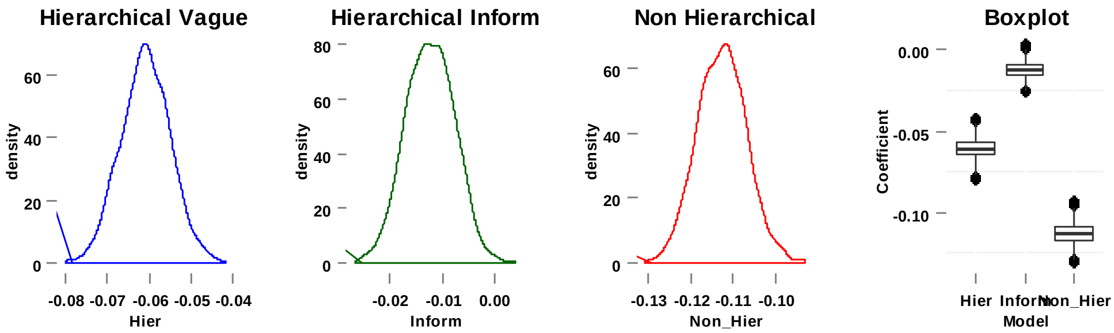

Figure 92: Supermarket own brand x Informational reinforcement (lower Utilitarian group) – beans

The estimates, in turn are -0.088, -0.012 and -0.113. The Bayesian posterior confidence intervals for the models are (-0.107, -0.070), (-0.025, 0.000) and (-0.129, -0.097) and t-statistics of -9.21 ($p<0.001$), -1.892 ($p=0.067$) and -13.914 ($p<0.001$). All three point estimates indicate a negative relationship between the informational reinforcement of the supermarket own brands on volume per purchase, though the Bayesian and frequentist inferences show this is statistically significant for the hierarchical vague and non-hierarchical models but not strictly for the hierarchical informative model. However, inspection of the confidence interval sees the value zero at the extremity and also the p-value of 0.067 is still significant at the 7% level and hence, with the strength of evidence from the other two models, it can be construed this variable is having a negative effect on volume per purchase. Therefore, it seems that the informational reinforcement of supermarket brands is having a negative effect on the volume per purchase. This is similar to the yellow fats category and again it can be hypothesised that consumers are not interested in informational reinforcement whilst shopping for supermarket own brands which are seen to have low utilitarian reinforcement value.

6.6.2.5 Supermarket own brand x Informational reinforcement in the higher Utilitarian reinforcement group (offset)

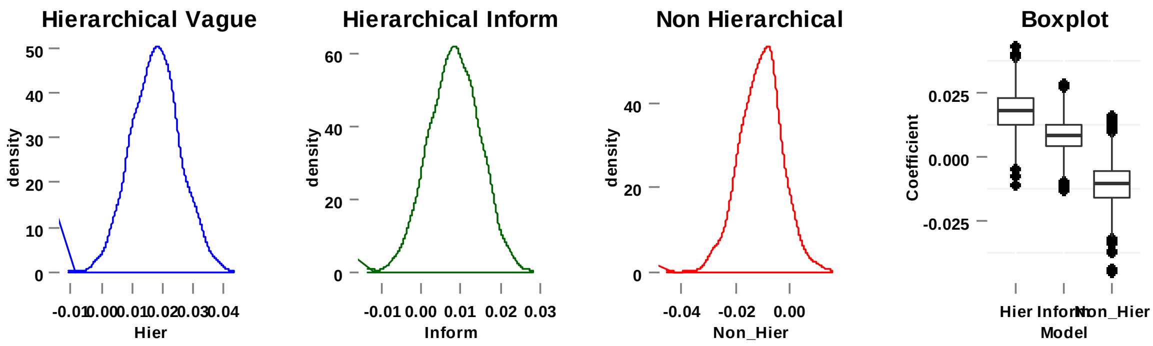

Figure 93: Supermarket own brand x Informational reinforcement (higher Utilitarian group) – beans

Fig 93 shows the density and box plots for the hierarchical vague, hierarchical informative and non-hierarchical models respectively.

The point estimates in the usual order are 0.031 with Bayesian confidence interval of (0.009, 0.054), t=2.713 (p=0.01) hence a positive significant relationship with volume per purchase. This is in contrast to the other two models which infer the parameter is no different from zero. This is seen from the hierarchical informed model with Bayesian confidence interval of (-0.009, 0.026), t=0.922 (p=0.261) and also the non-hierarchical model with Bayesian confidence intervals of (-0.033, 0.011), t= -0.955 (p=0.253). This again shines light on the importance of model structure and prior distribution definition. Therefore, the conclusion would be that the informational reinforcement of supermarket own brands within the higher utilitarian reinforcement group is, at best, having a positive effect on volume per purchase, above however two of the three models suggest this is not statistically significant.

6.6.2.6 Christmas effect x Informational reinforcement in the lower Utilitarian reinforcement group (offset)

The density and boxplots of the Christmas holiday week dummy variable can be seen graphically in Fig 94.

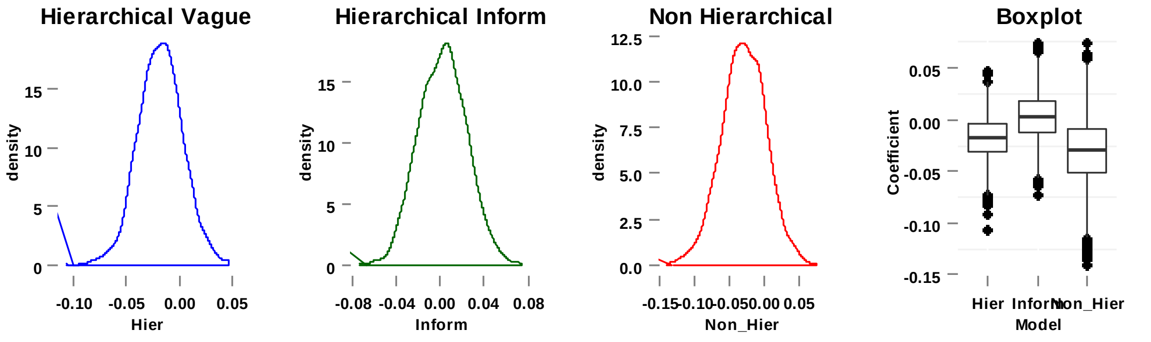

Figure 94: Christmas effect x Informational reinforcement (lower Utilitarian group) – beans

All models suggest the Christmas week for the lower Utilitarian reinforcement products has no effect on the volume per purchase within the beans category. This is due to the fact that Bayesian confidence intervals all straddle zero, (-0.063, 0.080), (-0.057, 0.063) and (-0.118, 058) and low value t-statistics, t=0.27 (p=0.385), t=0.085 (p=0.397), t= -0.654 (p=0.322) respectively. Hence the lower volume recognised in the category analysis section is due to fewer shopping days and, additionally less people visiting stores within that week.

6.6.2.7 Christmas effect x Informational reinforcement in the higher Utilitarian reinforcement group (offset)

Fig 95 shows the density and box plots for the Christmas week effect within the higher Utilitarian group. This is an offset measure to the effect within the lower utilitarian reinforcement group.

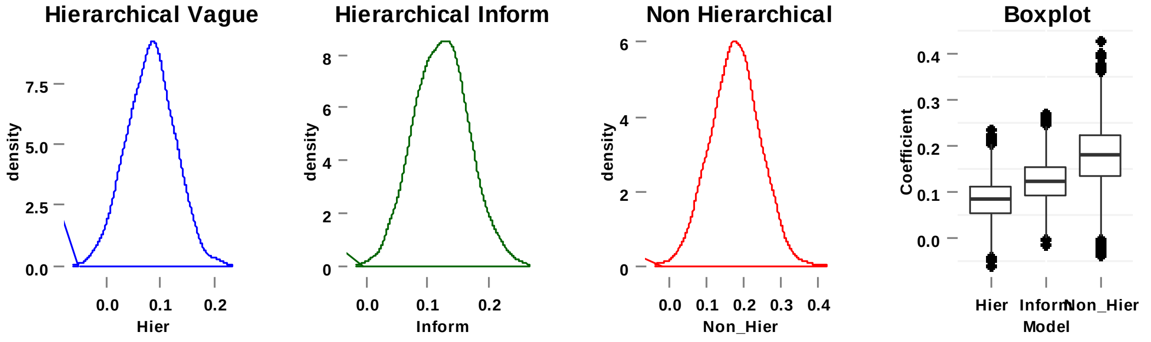

Figure 95: Christmas effect x Informational reinforcement (higher Utilitarian group) – beans

The hierarchical vague and non-hierarchical models infer no effect of this variable on the volume levels per purchase given their Bayesian confidence intervals, (-0.011, 0.277),

(-0.004, 0.365) and also t-statistics, t=1.78 (p=0.081) and t=1.90 (p=0.066), though it is worth noting these would suggest marginal evidence to recognising a positive effect. The hierarchical informative model would give differing conclusions between the Bayesian and frequentist measures (again under strict interpretation of the coefficients) whereby the 95% confidence interval does not include zero (0.001, 0.251). This means that the Bayesian would interpret this as a positive effect from this variable, whereby at a 95% level of confidence the frequentist would not reject the hypothesis this variable was zero given t=1.952 (p=0.059) and hence conclude the variable is not statistically significantly contributing to explaining the dependent volume variable. This again underlines the potential differences which are derived from different models structures and different prior distributions and, additionally in this case, different paradigm interpretations.

In reality it can be seen for all models there is a borderline result and the conclusion is there is weak evidence to suggest that during the Christmas week, there is a higher volume purchase being observed within the higher utilitarian reinforcement groups of products.

6.6.2.8 Characteristic variables.

Unlike the previous three categories, there is very limited evidence to suggest the product variants yield statistically different volumes per transaction with the variants having non-significant differences from the *beans only* base.

The number in pack variable is significant suggesting the larger packs have a larger volume per transaction than the single packs. There is a larger effect for 2+ pack size from the non-hierarchical structure model indicating the differences in interpretation given a hierarchical structure.

Chapter 7: Combined Category Model

7.1 Introduction

The study continues through the combining the four categories into one stacked data set as discussed in the methodology chapter. Given the homogeneity of the BPM variables and also a logged volume dependent and logged price independent variable, the model is valid to be run as one cross-category model. The non-focal variables are kept category specific as discussed in the methodology chapter.

The model is run with two functional forms, namely a pooled structure and a fixed effects offset structure. Within these functional forms, the model is run as a hierarchical and non-hierarchical structure. The models are estimated using a Gibbs sampler to produce the Bayesian MCMC with two chains. A burn-in of 4,000 iterations per chain is run to allow for parameter convergence and a further 2,000 iterations is used to assess model diagnostics and parameter inference of a Bayesian and frequentist nature. Results are discussed and a comparison of the models is offered.

Finally, a discussion is offered as to the comparison between the separate category models and the combined category model. Advantages and limitations regarding the combined category model are discussed bit statistically and theoretically.

7.2 Pooled Models

7.2.1 Model diagnostics

The standard deviations of the parameters are small and the posterior density plots of the coefficients show a relatively tight range, signifying convergence and are robustly normally distributed which is expected given the prior distribution assumptions.
The output of both hierarchical and non-hierarchical pooled models is displayed in Table 29. The same diagnostic statistics and parameter inference as the separate models are used.

	Pooled Non Hierarchical							Pooled Hierarchical						
	Beta	(SE posterior)	Bayes CI		t	sig		Beta	(SE posterior)	Bayes CI		t	sig	
Constant	4.491	0.010	4.472, 4.51	^	456.928	0.000	**	4.529	0.011	4.507, 4.55	^	412.603	0.000	**
Constant fj (offset vs. bis)	3.847	0.015	3.817, 3.875	^	265.262	0.000	**	3.837	0.014	3.809, 3.865	^	272.836	0.000	**
Constant yf (offset vs. bis)	3.500	0.014	3.472, 3.528	^	241.746	0.000	**	3.532	0.015	3.503, 3.559	^	241.561	0.000	**
Constant bb (offset vs. bis)	2.859	0.013	2.834, 2.884	^	220.329	0.000	**	2.928	0.013	2.903, 2.954	^	227.443	0.000	**
Log Price	-0.591	0.003	).597, -0.585	^	-204.782	0.000	**	-0.602	0.003	-0.607, -0.596	^	-208.005	0.000	**
Informational x Utilitarian Gp1	0.111	0.002	0.106, 0.116	^	44.995	0.000	**	0.097	0.002	0.092, 0.101	^	39.733	0.000	**
Informational x Utilitarian Gp2 vs. Gp	-0.006	0.004	-0.015, 0.002		-1.550	0.120		0.001	0.004	-0.007, 0.009		0.257	0.386	
SuperOwn x Informational	-0.033	0.003	).038, -0.028	^	-12.266	0.000	**	-0.036	0.003	-0.041, -0.031	^	-13.747	0.000	**
SuperOwn x Informational 2	-0.046	0.004	).055, -0.038	^	-11.060	0.000	**	-0.042	0.004	-0.049, -0.034	^	-10.605	0.000	**
Christmas	0.018	0.014	-0.011, 0.045		1.242	0.185		0.012	0.013	-0.013, 0.038		0.947	0.255	
Christmas Ut2	0.093	0.009	0.074, 0.111	^	10.103	0.000	**	0.082	0.009	0.065, 0.1	^	9.524	0.000	**
Chocolate Coated bis	0.166	0.007	0.152, 0.18	^	23.013	0.000	**	0.153	0.007	0.14, 0.166	^	22.860	0.000	**
Plain Sweet bis	0.182	0.008	0.166, 0.198	^	21.621	0.000	**	0.169	0.008	0.154, 0.185	^	21.365	0.000	**
Filled bis	0.024	0.009	0.007, 0.041	^	2.692	0.011	*	0.003	0.009	-0.014, 0.02		0.371	0.372	
Non Sweet bis	-0.019	0.009	).036, -0.002	^	-2.127	0.042	*	-0.030	0.008	-0.047, -0.014	^	-3.582	0.001	**
Countlines bis	base							base						
Size 2-5 bis	0.189	0.009	0.172, 0.206	^	21.193	0.000	**	0.181	0.008	0.165, 0.197	^	22.184	0.000	**
Size 6-7 bis	0.082	0.007	0.068, 0.096	^	11.516	0.000	**	0.091	0.007	0.077, 0.105	^	12.939	0.000	**
Size 8-11 bis	0.194	0.008	0.178, 0.21	^	23.451	0.000	**	0.184	0.008	0.168, 0.199	^	23.578	0.000	**
Size 12+ bis	0.374	0.007	0.359, 0.388	^	51.340	0.000	**	0.340	0.007	0.326, 0.353	^	48.091	0.000	**
Size packs bis	0.571	0.010	0.551, 0.591	^	54.634	0.000	**	0.554	0.010	0.536, 0.573	^	57.922	0.000	**
Size 1s bis	base							base						
Other fruit fj	0.033	0.034	-0.036, 0.101		0.954	0.253		0.093	0.033	0.028, 0.159	^	2.803	0.008	**
Breakfast fj	-0.068	0.049	-0.165, 0.027		-1.375	0.155		-0.050	0.046	-0.14, 0.04		-1.067	0.226	
Grape fj	0.152	0.021	0.111, 0.192	^	7.384	0.000	**	0.161	0.019	0.122, 0.199	^	8.271	0.000	**
Grapefruit fj	-0.120	0.017	).154, -0.087	^	-6.940	0.000	**	-0.037	0.017	-0.07, -0.005	^	-2.251	0.032	*
Mixed fj	-0.016	0.015	-0.044, 0.014		-1.055	0.229		0.011	0.014	-0.017, 0.038		0.801	0.289	
Orange fj	0.013	0.009	-0.004, 0.032		1.455	0.138		0.035	0.009	0.017, 0.052	^	3.930	0.000	**
Pineapple fj	-0.263	0.016	).295, -0.232	^	-16.259	0.000	**	-0.185	0.015	-0.216, -0.156	^	-12.100	0.000	**
Tomato fj	-0.295	0.025	).344, -0.247	^	-11.809	0.000	**	-0.215	0.024	-0.264, -0.167	^	-8.885	0.000	**
Vegetable fj	0.074	0.056	-0.034, 0.184		1.317	0.168		0.044	0.053	-0.057, 0.148		0.839	0.281	
Vitamin fj	-0.017	0.079	-0.171, 0.139		-0.210	0.390		-0.008	0.073	-0.152, 0.135		-0.104	0.397	
Apple fj	base							base						
size 2-5 fj	0.336	0.011	0.313, 0.358	^	29.354	0.000	**	0.347	0.011	0.326, 0.368	^	32.482	0.000	**
Size 6+ fj	0.647	0.025	0.597, 0.696	^	25.437	0.000	**	0.601	0.025	0.552, 0.649	^	24.490	0.000	**
Size 1s fj	base							base						
Butter yf	-0.250	0.007	).263, -0.235	^	-34.629	0.000	**	-0.259	0.007	-0.273, -0.245	^	-36.214	0.000	**
Margarine yf	-0.189	0.008	).205, -0.174	^	-23.673	0.000	**	-0.192	0.008	-0.208, -0.177	^	-24.171	0.000	**
Low Reduced yf	-0.110	0.009	).127, -0.093	^	-12.691	0.000	**	-0.122	0.009	-0.139, -0.105	^	-14.334	0.000	**
Blended spreads yf	base							base						
Size 2+ yf	0.291	0.041	0.211, 0.374	^	7.081	0.000	**	0.313	0.040	0.235, 0.389	^	7.875	0.000	**
Size 1s yf	base							base						
Beans Plus yf	0.001	0.012	-0.022, 0.024		0.104	0.397		-0.020	0.011	-0.043, 0.002		-1.771	0.083	
Tomato bb	-0.009	0.011	-0.03, 0.012		-0.806	0.288		-0.011	0.010	-0.032, 0.009		-1.081	0.223	
Healthy bb	-0.009	0.019	-0.045, 0.028		-0.453	0.360		-0.015	0.017	-0.049, 0.019		-0.881	0.271	
Flavours bb	-0.054	0.030	-0.114, 0.008		-1.772	0.083		-0.047	0.028	-0.099, 0.008		-1.684	0.097	
Beans Only bb	base							base						
Size 4-12 bb	1.074	0.010	1.054, 1.094	^	104.818	0.000	**	0.997	0.010	0.977, 1.016	^	98.988	0.000	**
Size 1-2 bb	base							base						

	Pooled Non Hierarchical	Pooled Hierarchical
R-Squared (adj)	69.447%	72.202%
Mean Deviance	179,127	154,772
Penalty	42	1,564
DIC	179,169	156,336
MAPE	6.636%	6.207%
Variance (between purchases)	0.093	0.082
Variance (between housholds)		0.153
Variance (between household) t-stat (sig)		22.86(0)
Variance Partition Coeficient		65.043%

* significant 5%
** significant 1%
^ 95% Bayesian estimates do not include zero

Table 29: Model diagnostics and inference - pooled

From the Bayesian inference, it can be seen the mean deviance of the hierarchical model and non-hierarchical models are 154,772 and 179,127 respectively with a penalty of 1,564 and 42, resulting in a DIC of 156,336 and 179,169 respectively. Hence the hierarchical model is proving a better representation of the data despite the increased penalty for a more complicated model (Spiegelhalter *et al.*, 2002). The R-squared (adj) value for the hierarchical model and the non-hierarchical model is 72.202% and 69.447% respectively indicating the hierarchical model is a better fit to the data, even taking the more complex structure into account (Field *et al.*, 2012). The MAPE is also smaller for the hierarchical model (6.210% vs. 6.643% for the non-hierarchical) indicating that the average absolute error of the hierarchical model is smaller. The variance of the hierarchical model is of a smaller magnitude than the non-hierarchical model (0.198 vs 0.240) implying the model is accounting for a larger proportion of the variance of the data. The hierarchical variance term of 0.153 has a small standard error resulting in a highly significantly large t-statistic (58.36) which rejects the null hypothesis that this parameter is equal to zero ($p<0.001$). This variance produces a variance partition coefficient of 65.403%, suggesting the hierarchical structure is an important element of the model.

Both models offer a good representation of the underlying data though the hierarchical pooled model diagnostics suggest this model is preferred to the non-hierarchical pooled model, at least statistically.

7.2.2 Coefficient discussion

7.2.2.1 Price elasticity

Fig 96 shows the density and box plots for the posterior distribution of the pooled hierarchical and pooled non-hierarchical models. The point estimates are -0.602 and -0.591 respectively.

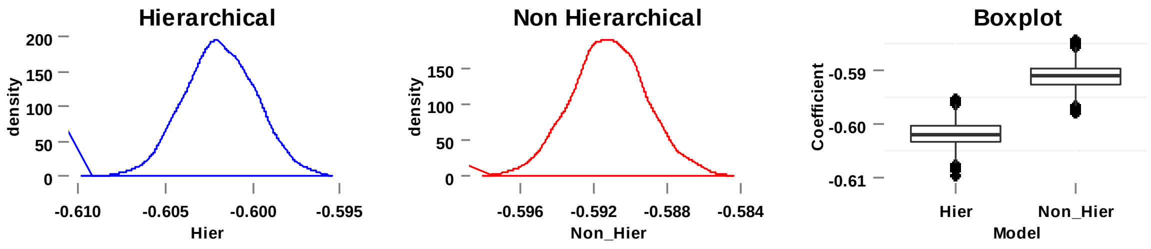

Figure 96: Price coefficients - pooled

The price coefficient for both the hierarchical and non-hierarchical models are statistically different from zero given the Bayesian posterior confidence intervals of (-0.607, -0.596) and (-0.597, -0.585) and t-statistics of -208. 0 and -204.8 respectively and hence rejects the hypothesis these values are zero with $p<0.001$. The non-overlapping nature of the boxplot would also suggest this to be the case.

7.2.2.2 Informational reinforcement in the lower Utilitarian reinforcement group

Fig 97 shows the informational variable of the lower utilitarian group for both models plotted as a density plot and as a box plot comparing both model estimates. There is graphical evidence from the plots the statistics are significantly different.

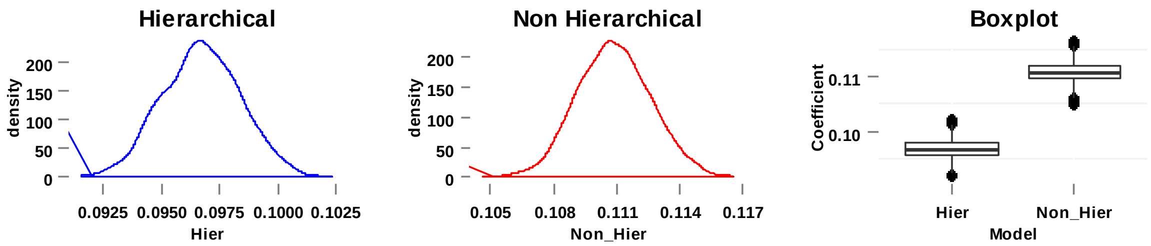

Figure 97: Informational reinforcement (lower Utilitarian group) coefficients - pooled

The estimate for the hierarchical and non-hierarchical models respectively are 0.097 and 0.111 with Bayesian confidence intervals of (0.092, 0.101) and (0.106, 0.116) and t-statistics of 39.7 and 45.0 (both $p<0.001$), hence strong evidence from both a Bayesian and frequentist perspective to suggest the parameter is non-zero and positive. In each model's case, across the four categories, the informational reinforcement in the lower utilitarian group is positively influencing volume per purchase above and beyond what can be accounted for by price.

7.2.2.3 Informational reinforcement in the higher Utilitarian reinforcement group (offset)

Fig 98 shows the density and box plots offset value for the informational variable in the higher utilitarian group (offset against the base informational value) for the hierarchical and non-hierarchical models.

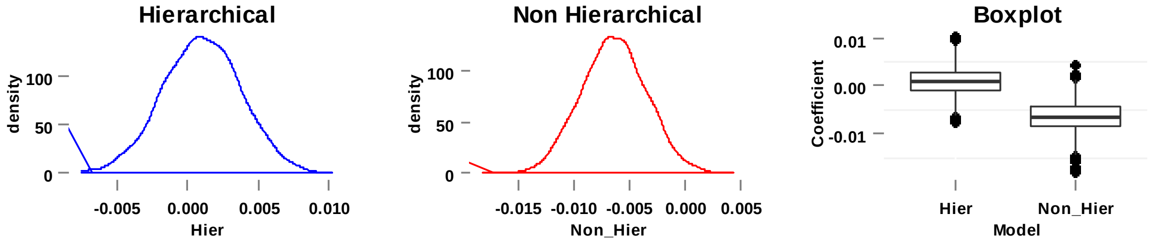

Figure 98: Informational reinforcement (higher Utilitarian group) coefficients - pooled

The point estimate of the offset for each respective model is 0.001 and -0.006 for the hierarchical and non-hierarchical models. The Bayesian confidence intervals are (-0.007, 0.009) and (-0.015, 0.002) and t-statistics of 0.257 (p=0.386) and -1.55 (p=0.12) respectively leading to the conclusion the parameters are zero under both the Bayesian and frequentist paradigms. Therefore, there is no further effect from informational reinforcement group within the higher utilitarian reinforcement group, above and beyond what is reinforced from the lower utilitarian reinforcement group.

7.2.2.4 Supermarket own brand x Informational reinforcement in the lower Utilitarian reinforcement group

Fig 99 shows the density and box plots of the informational variable of supermarket own brands estimates of the hierarchical and non-hierarchical models. The point estimates are -0.036 and -0.033.

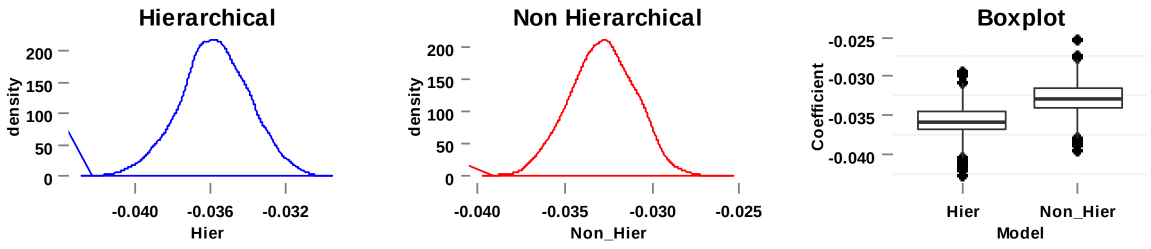

Figure 99: Supermarket own brand x Informational reinforcement (lower Utilitarian group) - pooled

The informational variable crossed with the supermarket own indicator is negative and statistically significant, given the Bayesian confidence intervals (-0.041, -0.031) and (-0.038 and -0.028) respectively. The intervals overlap suggesting the estimates of both models are statistically similar. Neither interval contains zero suggesting they are statistically important to the model. Also, the frequentist t-statistics of -13.747 and -12.266 respectively, both $p<0.001$, show strong evidence the parameters are statistically significant. This would suggest a negative impact of volume would be seen for supermarket own brands with increased informational reinforcement within the lower utilitarian reinforcement group. This suggests as consumers shop for supermarket own brands within a lower utilitarian reinforcement group, products showing higher informational reinforcements are less appealing.

7.2.2.5 Supermarket own brand x Informational reinforcement in the higher Utilitarian reinforcement group (offset)

Fig 100 shows the density and box plots for the hierarchical and non-hierarchical posterior distribution of the parameter. The point estimates, in turn, are -0.042 and -0046.

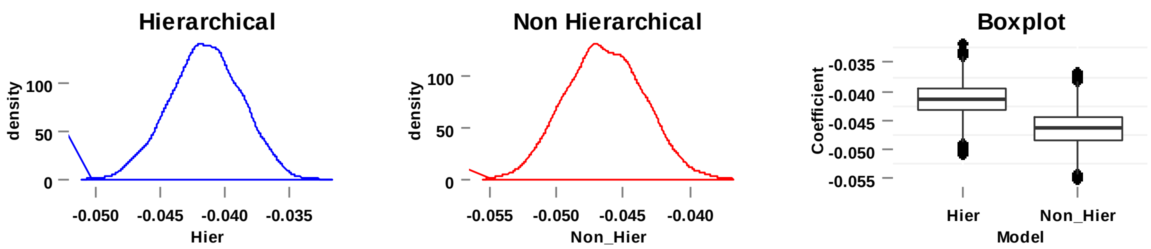

Figure 100: Supermarket own brand x Informational reinforcement (higher Utilitarian group) – pooled

The Bayesian posterior confidence intervals are (-0.049, -0.034) for the hierarchical and (-0.055, -0.038) for the non-hierarchical models, neither interval containing the value zero. The t-statistics of the models, in the usual order, are -10.605 and -11.060, both $p<0.001$, hence strong evidence the parameter is non-zero. Given this is an offset to the effect within the lower utilitarian group, it implies the volume is adversely affected above and beyond what is observed in the lower utilitarian reinforcement group. Hence volume per purchase is negatively affected for supermarket own brands with higher levels of informational and higher utilitarian reinforcement. Furthermore, this negative effect is stronger within the higher utilitarian reinforcement group than the lower group. It would seem consumers are seeking utilitarian reinforcement from supermarket own brands rather than the informational reinforcement of the products, at least whilst analysing the results in a pooled model structure.

7.2.2.6 Christmas effect x Informational reinforcement in the lower Utilitarian reinforcement group

The posterior distribution of the effect of the Christmas week on the lower utilitarian reinforcement group is shown graphically in Fig 101 and the point estimates for the two models, given in the usual order are 0.012 and 0.018.

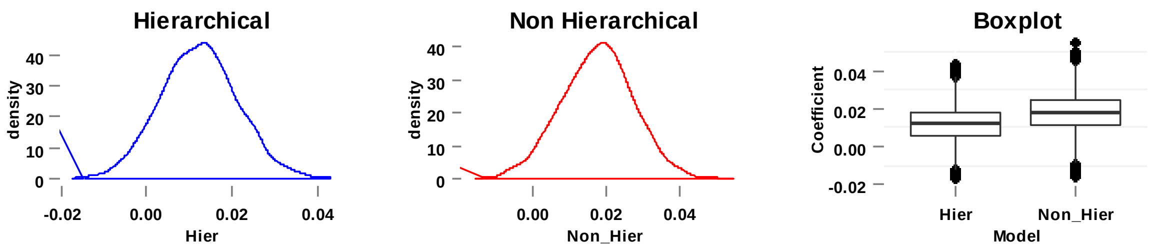

Figure 101: Christmas effect x Informational reinforcement (lower Utilitarian group) - pooled

The Bayesian posterior confidence intervals are (-0.013, 0.038) and (-0.011, 0.045) respectively, both straddling the value zero. The small t-statistics and non-significant p-values (t=0.947, p=0.255 and t=1.242, p=0.185 respectively) also indicate these are non-significant and hence close to zero. This implies the Christmas week is no different from other weeks in terms of volume per purchase within the lower utilitarian reinforcement group.

7.2.2.6 Christmas effect x Informational reinforcement in the higher Utilitarian reinforcement group (offset)

Fig 102 shows the density and box plots for the Christmas week effect within the higher Utilitarian group, as an offset to the Christmas week x Informational Reinforcement within the lower Utilitarian group.

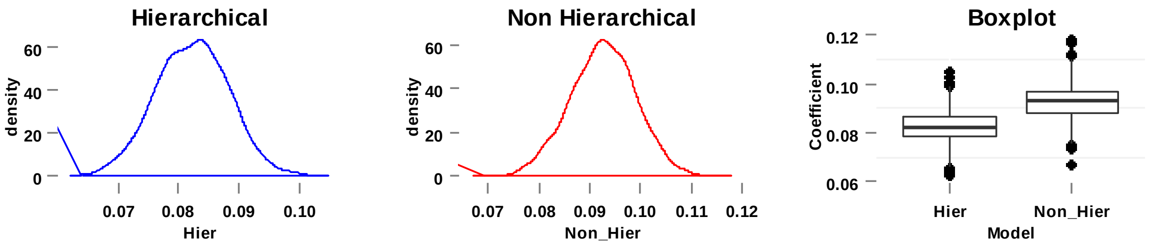

Figure 102: Christmas effect x Informational reinforcement (higher Utilitarian group) - pooled

The point estimates for the posterior distributions of the hierarchical and non-hierarchical models are 0.082 and 0.093 respectively, with Bayesian confidence intervals of (0.065, 0.100) and (0.74, 0.111). This suggests the probability of the parameter being positive is high. The frequentist t-statistics of 9.5 and 10.1, both $p<0.001$ also indicate the parameter is statistically significantly positive. This means that compared to the lower Utilitarian reinforcement group, the higher Utilitarian reinforcement group of brands have a higher volume per purchase in the Christmas week compared to the average week.
This implication suggests consumers are purchasing more volume during the Christmas week from the higher Utilitarian reinforcement group but the lower utilitarian reinforcement group sees no statistical difference from an average week in terms of volume purchased. This effect is above and beyond what can be explained by price changes, informational/utilitarian reinforcement categorisation and supermarket own brand effect.

7.3 Summary of Pooled Models

The pooled model approach by combining the four categories has resulted in the development of hierarchical and non-hierarchical models, both of which give a diagnostically good representation of the underlying data. A comparison of the parameters with the other models will be discussed in more detail when all models are accounted for, however there are conclusions emerging from the data which suggest the practitioner can take insights from the economic variables and the BPM variables when considering the behaviour of consumers within this defined four category marketplace.

The nature of the underlying structure of the model, whether the underlying hierarchy of the data is taken into account again has an impact on the parameter assessment of the model. This was also seen in the separate category analysis, strengthening the argument that the assumptions made in model creation impacts the values of the parameters. This is in line with (Leamer, 1992; Rossi and Allenby, 2003; Gelman, 2010) who stated that the nature of the model structure will impact model output, not only from a Bayesian prior probability distribution point of view but also the structure of the model which would be equally as relevant to a frequentist approach.

This is discussed in more detail following the analysis of the fixed effect models next.

7.3 Fixed Effects Models

7.3.1 Model diagnostics

The next section also focuses on the combined category model, though this time the model is run as a fixed effect model, as to a pooled structure, as discussed in the methods chapter. As with the pooled model, the model is calculated using the Bayesian MCMC Gibbs sampler with two chains. A burn-in of 4,000 iterations per chain and a further 2,000 are used to estimate the parameter inference. Model diagnostics and parameter inference are calculated using both Bayesian and frequentist methods. The model structure utilises a hierarchical and non-hierarchical functional form.

Table 30 shows the diagnostics of the model together with the coefficients.

	Fixed Effects Non Hierarchical						Fixed Effects Hierarchical					
	Beta (SE posterior)		Bayes CI		t	sig	Beta (SE posterior)		Bayes CI		t	sig
Constant bis	4.489	0.010	4.47, 4.508	^	462.654	0.000 **	4.538	0.011	4.516, 4.559	^	422.648	0.000 **
Constant fj (offset vs. bis)	3.559	0.019	3.522, 3.597	^	187.456	0.000 **	3.601	0.019	3.563, 3.639	^	185.562	0.000 **
Constant yf (offset vs. bis)	3.041	0.024	2.994, 3.086	^	129.079	0.000 **	3.044	0.023	2.999, 3.09	^	131.111	0.000 **
Constant bb (offset vs. bis)	3.053	0.022	3.01, 3.096	^	139.951	0.000 **	3.056	0.021	3.015, 3.097	^	146.783	0.000 **
Log Price bis	-0.701	0.004	-0.709, -0.693	^	-171.981	0.000 **	-0.696	0.004	-0.703, -0.688	^	-178.163	0.000 **
Log Price fj (offset vs. bis)	0.208	0.008	0.192, 0.224	^	24.682	0.000 **	0.171	0.008	0.155, 0.188	^	20.523	0.000 **
Log Price yf (offset vs. bis)	0.246	0.008	0.229, 0.262	^	29.669	0.000 **	0.238	0.008	0.222, 0.253	^	29.961	0.000 **
Log Price bb (offset vs. bis)	0.131	0.012	0.108, 0.155	^	11.085	0.000 **	0.125	0.011	0.103, 0.148	^	11.057	0.000 **
Informational x Utilitarian Gp1 bis	0.027	0.003	0.02, 0.034	^	7.851	0.000 **	0.032	0.003	0.025, 0.038	^	9.619	0.000 **
Informational x Utilitarian Gp1 fj (offset vs. bis)	0.172	0.008	0.157, 0.189	^	21.167	0.000 **	0.129	0.008	0.114, 0.145	^	16.575	0.000 **
Informational x Utilitarian Gp1 yf (offset vs. bis)	0.146	0.006	0.133, 0.158	^	22.992	0.000 **	0.116	0.006	0.104, 0.128	^	18.690	0.000 **
Informational x Utilitarian Gp1 bb (offset vs. bis)	-0.025	0.008	-0.04, -0.01	^	-3.299	0.002 **	-0.024	0.007	-0.038, -0.01	^	-3.342	0.001 **
Informational x Utilitarian Gp2 vs. Gp1 bis	0.073	0.004	0.065, 0.081	^	17.054	0.000 **	0.058	0.004	0.05, 0.066	^	14.264	0.000 **
Informational x Utilitarian Gp2 vs. Gp1 fj (offset vs. bis)	-0.034	0.013	-0.06, -0.007	^	-2.521	0.017 *	-0.033	0.013	-0.059, -0.007	^	-2.464	0.019 *
Informational x Utilitarian Gp2 vs. Gp1 yf (offset vs. bis)	-0.091	0.006	-0.104, -0.079	^	-14.544	0.000 **	-0.060	0.006	-0.071, -0.048	^	-10.248	0.000 **
Informational x Utilitarian Gp2 vs. Gp1 bb (offset vs. bis)	-0.032	0.008	-0.047, -0.017	^	-4.253	0.000 **	-0.009	0.007	-0.023, 0.005		-1.221	0.189
SuperOwn x Informational bis	0.007	0.004	0, 0.014		1.831	0.075	-0.001	0.004	-0.008, 0.006		-0.367	0.373
SuperOwn x Informational fj (offset vs. bis)	-0.041	0.007	-0.055, -0.026	^	-5.591	0.000 **	-0.032	0.007	-0.046, -0.019	^	-4.738	0.000 **
SuperOwn x Informational yf (offset vs. bis)	-0.064	0.008	-0.079, -0.049	^	-8.508	0.000 **	-0.042	0.007	-0.056, -0.028	^	-5.826	0.000 **
SuperOwn x Informational bb (offset vs. bis)	-0.120	0.009	-0.138, -0.102	^	-13.551	0.000 **	-0.121	0.008	-0.138, -0.105	^	-14.384	0.000 **
SuperOwn x Informational2 bis	-0.092	0.005	-0.103, -0.082	^	-17.755	0.000 **	-0.080	0.005	-0.089, -0.07	^	-16.015	0.000 **
SuperOwn x Informational2 fj (offset vs. bis)	0.002	0.016	-0.029, 0.034		0.153	0.394	0.037	0.016	0.007, 0.068	^	2.338	0.026 *
SuperOwn x Informational2 yf (offset vs. bis)	0.144	0.013	0.119, 0.169	^	11.302	0.000 **	0.102	0.012	0.078, 0.125	^	8.340	0.000 **
SuperOwn x Informational2 bb (offset vs. bis)	0.082	0.012	0.058, 0.105	^	6.737	0.000 **	0.061	0.011	0.038, 0.083	^	5.316	0.000 **
Christmas bis	0.058	0.030	0.001, 0.116	^	1.962	0.058	0.059	0.027	0.007, 0.113	^	2.192	0.036 *
Christmas fj (offset vs. bis)	-0.054	0.045	-0.144, 0.034		-1.182	0.198	-0.045	0.041	-0.125, 0.036		-1.102	0.217
Christmas yf (offset vs. bis)	-0.122	0.041	-0.202, -0.04	^	-2.950	0.005 **	-0.126	0.037	-0.199, -0.054	^	-3.376	0.001 **
Christmas bb (offset vs. bis)	-0.087	0.055	-0.195, 0.017		-1.592	0.112	-0.067	0.049	-0.166, 0.03		-1.366	0.157
Christmas Ut2 bis	0.008	0.044	-0.076, 0.092		0.174	0.393	-0.018	0.040	-0.097, 0.06		-0.434	0.363
Christmas Ut2 fj (offset vs. bis)	0.069	0.106	-0.136, 0.28		0.651	0.323	0.097	0.094	-0.091, 0.283		1.028	0.235
Christmas Ut2 yf (offset vs. bis)	0.064	0.080	-0.093, 0.218		0.807	0.288	0.050	0.075	-0.095, 0.197		0.675	0.318
Christmas Ut2 bb (offset vs. bis)	0.170	0.104	-0.033, 0.373		1.630	0.106	0.176	0.095	-0.007, 0.37		1.859	0.071
Chocolate Coated bis	0.152	0.007	0.139, 0.166	^	21.737	0.000 **	0.144	0.007	0.131, 0.157	^	21.792	0.000 **
Plain Sweet bis	0.160	0.010	0.141, 0.179	^	16.800	0.000 **	0.132	0.009	0.115, 0.15	^	14.679	0.000 **
Filled bis	-0.011	0.009	-0.029, 0.007		-1.278	0.176	-0.026	0.008	-0.043, -0.01	^	-3.126	0.003 **
Non Sweet bis	0.039	0.011	0.018, 0.059	^	3.657	0.000 **	-0.004	0.010	-0.024, 0.015		-0.440	0.362
Countlines bis	base						base		0.193, 0.196			**
Size 2-5 bis	0.206	0.009	0.19, 0.223	^	23.924	0.000 **	0.194	0.008	0.179, 0.21	^	24.230	0.000 **
Size 6-7 bis	0.086	0.007	0.072, 0.101	^	11.594	0.000 **	0.092	0.007	0.079, 0.106	^	13.447	0.000 **
Size 8-11 bis	0.194	0.008	0.178, 0.21	^	23.404	0.000 **	0.183	0.008	0.168, 0.198	^	24.111	0.000 **
Size 12+ bis	0.360	0.007	0.346, 0.374	^	49.137	0.000 **	0.327	0.007	0.314, 0.341	^	46.967	0.000 **
Size packs bis	0.590	0.010	0.57, 0.61	^	57.825	0.000 **	0.569	0.010	0.55, 0.588	^	58.573	0.000 **
Size 1s bis	base						base					**
Other fruit fj	0.014	0.035	-0.057, 0.083		0.388	0.370	0.071	0.033	0.006, 0.138	^	2.129	0.041 *
Breakfast fj	-0.166	0.049	-0.261, -0.07	^	-3.422	0.001 **	-0.123	0.046	-0.21, -0.031	^	-2.691	0.011 *
Grape fj	0.107	0.021	0.066, 0.148	^	5.111	0.000 **	0.131	0.020	0.091, 0.169	^	6.595	0.000 **
Grapefruit fj	-0.155	0.017	-0.188, -0.121	^	-9.097	0.000 **	-0.059	0.017	-0.092, -0.026	^	-3.500	0.001 **
Mixed fj	-0.028	0.015	-0.057, 0.002		-1.804	0.078	0.006	0.015	-0.023, 0.034		0.427	0.364
Orange fj	0.013	0.009	-0.005, 0.031		1.472	0.135	0.038	0.009	0.02, 0.055	^	4.201	0.000 **
Pineapple fj	-0.275	0.016	-0.305, -0.243	^	-17.088	0.000 **	-0.195	0.016	-0.225, -0.165	^	-12.575	0.000 **
Tomato fj	-0.311	0.025	-0.359, -0.262	^	-12.464	0.000 **	-0.229	0.024	-0.275, -0.184	^	-9.601	0.000 **
Vegetable fj	0.114	0.058	-0.002, 0.229		1.971	0.057	0.082	0.053	-0.021, 0.188		1.532	0.123
Vitamin fj	-0.141	0.081	-0.302, 0.015		-1.750	0.086	-0.070	0.074	-0.218, 0.072		-0.939	0.257
Apple fj	base						base					**
size 2-5 fj	0.325	0.011	0.304, 0.347	^	28.709	0.000 **	0.336	0.011	0.314, 0.358	^	30.378	0.000 **
Size 6+ fj	0.589	0.025	0.539, 0.641	^	23.140	0.000 **	0.560	0.025	0.51, 0.608	^	22.513	0.000 **
Size 1s fj	base						base					**
Butter yf	-0.308	0.008	-0.323, -0.292	^	-38.569	0.000 **	-0.324	0.008	-0.339, -0.308	^	-40.717	0.000 **
Margarine yf	-0.187	0.008	-0.202, -0.171	^	-23.001	0.000 **	-0.181	0.008	-0.196, -0.165	^	-22.906	0.000 **
Low Reduced yf	-0.121	0.009	-0.139, -0.103	^	-13.336	0.000 **	-0.125	0.009	-0.142, -0.108	^	-14.268	0.000 **
Blended spreads yf	base						base					**
Size 2+ yf	0.427	0.042	0.344, 0.508	^	10.194	0.000 **	0.442	0.039	0.366, 0.517	^	11.486	0.000 **
Size 1s yf	base						base					**
Beans Plus yf	0.004	0.012	-0.02, 0.028		0.335	0.377	-0.016	0.011	-0.038, 0.007		-1.387	0.152
Tomato bb	-0.007	0.011	-0.029, 0.014		-0.650	0.323	-0.009	0.010	-0.029, 0.012		-0.863	0.275
Healthy bb	-0.015	0.019	-0.051, 0.021		-0.816	0.286	-0.021	0.018	-0.056, 0.014		-1.192	0.196
Flavours bb	-0.052	0.030	-0.11, 0.006		-1.752	0.086	-0.048	0.027	-0.1, 0.006		-1.748	0.087
Beans Only bb	base						base					**
Size 4-12 bb	1.132	0.011	1.111, 1.154	^	101.973	0.000 **	1.048	0.011	1.027, 1.069	^	97.605	0.000 **
Size 1-2 bb	base						base					
R-Squared (adj)	70.435%						72.858%					
Mean Deviance	176,177						152,502					
Penalty	63						1,584					
DIC	176,240						154,087					
MAPE	6.534%						6.126%					
Variance (between purchases)	0.235						0.195					
Variance (between housholds)							0.045					
Variance (between household) t-stat (sig)							57.881(0)					
Variance Partition Coeficient							18.766%					

* significant 5%

** significant 1%

^ 95% Bayesian estimates do not include zero

Table 30: Model diagnostics and inference - fixed effect

From Table 30 it can be seen the mean deviance for the hierarchical model is 152,502 with a penalty of 1,584 resulting in a DIC of 154,087. This is compared to the same statistics for the non-hierarchical model of 176,177, 63 and 176,240, therefore the hierarchical model is better representing a data set of a similar structure despite the increased penalty due to the more complex model functional form (Spiegelhalter *et al.*, 2002). The R-squared (adjusted) for each model in turn is 72.858% and 70.435% which favours the hierarchical model a little over the non-hierarchical. The MAPE shows the average error per observation is lower for the hierarchical model (6.126%) than the non-hierarchical model (6.534%). Also, the total variance of the hierarchical model is lower (0.195) that that of the non-hierarchical model (0.235). Dividing the hierarchical variance estimate by its standard error gives a t-statistic of 57.881 ($p<0.001$) which suggests strong evidence to reject the null hypothesis and that the parameter is redundant. This results in a variance partition coefficient of 18.766%. The diagnostics suggests the hierarchical model is preferred statistically to the non-hierarchical model.

7.3.2 Coefficient discussion

Attention is now turned to the estimates of the coefficients. There is a base coefficient (corresponding to the biscuit category) and offsets which reflect the deviance from the biscuit category and hence the inferential statistics relate to this offset over and above the estimate of the base (biscuit) category. Each coefficient will include a table showing the base biscuit point estimate of the parameter along with the Bayesian confidence intervals of the posterior distribution and the frequentist t-statistic and significance level. The other categories will include the same statistics however it will represent a deviance from the base biscuit category and hence it will be able to judge whether each category is statistically similar to the biscuit category coefficient or not.

The actual value of the coefficients for each category (rather than the offsets) are also formulated, as described in the methods section.

7.3.2.1 Price elasticity

Price	Non-Hierarchical						Hierarchical					
	Estimate	Bayes CI	t-stat	sig	Constructed Est & CI		Estimate	Bayes CI	t-stat	sig	Constructed Est & CI	
Biscuits	-0.701	-0.709, -0.693	-171.98	0.000	-0.701	-0.709, -0.693	-0.696	-0.703, -0.688	-178.16	0.000	-0.696	-0.703, -0.688
Fruit Juice Offset to Biscuits	0.208	0.192, 0.224	24.68	0.000	-0.493	-0.499, -0.488	0.171	0.155, 0.188	20.52	0.000	-0.524	-0.53, -0.519
Yellow Fat Offset to Biscuits	0.246	0.229, 0.262	29.67	0.000	-0.456	-0.462, -0.45	0.238	0.222, 0.253	29.96	0.000	-0.458	-0.463, -0.452
Baked Beans Offset to Biscuits	0.131	0.108, 0.155	11.09	0.000	-0.570	-0.58, -0.561	0.125	0.103, 0.148	11.06	0.000	-0.570	-0.578, -0.563

Table 31: Price coefficients - offset fixed effect

From Table 31 it can be seen the point estimate of the coefficient for the base category of biscuits is similar for both models (-0.701 hierarchical and -0.696 non-hierarchical) with Bayesian confidence intervals which overlap suggesting they are statistically similar. The confidence intervals do not include zero and the t-statistics are significant at $p<0.001$. This implies a negative elasticity measure, similar to both models. The measure is in line with the separate models (discussed later) and other studies involving this category[4].

All other categories have a positive offset to the biscuit category and this is the case for both hierarchical and non-hierarchical models. These offsets are also statistically relevant given the Bayesian confidence intervals indicate a very low probability the parameter is zero and also the high t-statistics (all $p<0.001$) indicating the parameter is statistically significantly different from zero and hence positive given the t-statistic sign. Hence the elasticity of demand for these categories is lower than the biscuit category, though all are similar and in line with other studies and other previous models within this study.

7.3.2.2 Informational reinforcement in the lower Utilitarian reinforcement group

Informational x Utilitarian Gp1	Non-Hierarchical						Hierarchical					
	Estimate	Bayes CI	t-stat	sig	Constructed Est & CI		Estimate	Bayes CI	t-stat	sig	Constructed Est & CI	
Biscuits	0.027	0.02, 0.034	7.85	0.000	0.027	0.02, 0.034	0.032	0.025, 0.038	9.62	0.000	0.032	0.025, 0.038
Fruit Juice Offset to Biscuits	0.172	0.157, 0.189	21.17	0.000	0.200	0.194, 0.205	0.129	0.114, 0.145	16.58	0.000	0.161	0.156, 0.166
Yellow Fat Offset to Biscuits	0.146	0.133, 0.158	22.99	0.000	0.173	0.169, 0.178	0.116	0.104, 0.128	18.69	0.000	0.148	0.144, 0.153
Baked Beans Offset to Biscuits	-0.025	-0.04, -0.01	-3.30	0.002	0.002	-0.003, 0.007	-0.024	-0.038, -0.01	-3.34	0.001	0.008	0.003, 0.013

Table 32: Informational reinforcement (lower Utilitarian group) coefficients - fixed effect

The inference of the coefficients of the informational reinforcement variable within the lower utilitarian reinforcement group for the biscuit (base) category are displayed in Table 32. The Bayesian confidence intervals and t-statistics indicate the estimates are statistically valid as model predictors given lack of the value zero within the confidence intervals and the high t-

[4] Non-hierarchical studies

statistics (all p<=0.002). The reconstructed estimates show differences between category and these differences are prevalent within the hierarchical and non-hierarchical models. The extremities are the baked beans at the lower end and the fruit juice at the higher. The hierarchical structure has resulted in some shrinkage of the parameter with all four estimates having a smaller variance between them.

7.3.2.3 Informational reinforcement in the higher Utilitarian reinforcement group

Informational x Utilitarian Gp2	Non-Hierarchical						Hierarchical					
	Estimate	Bayes CI	t-stat	sig	Constructed Est & CI		Estimate	Bayes CI	t-stat	sig	Constructed Est & CI	
Biscuits	0.073	0.065, 0.081	17.05	0.000	0.073	0.065, 0.081	0.058	0.05, 0.066	14.26	0.000	0.058	0.05, 0.066
Fruit Juice Offset to Biscuits	-0.034	-0.06, -0.007	-2.52	0.017	0.039	0.032, 0.047	-0.033	-0.059, -0.007	-2.46	0.019	0.025	0.017, 0.032
Yellow Fat Offset to Biscuits	-0.091	-0.104, -0.079	-14.54	0.000	-0.018	-0.023, -0.013	-0.060	-0.071, -0.048	-10.25	0.000	-0.002	-0.007, 0.003
Baked Beans Offset to Biscuits	-0.032	-0.047, -0.017	-4.25	0.000	0.041	0.035, 0.046	-0.009	-0.023, 0.005	-1.22	0.189	0.049	0.044, 0.054

Table 33: Informational reinforcement (higher Utilitarian group) coefficients – fixed effect

When considering the Informational reinforcement for the higher Utilitarian reinforcement group (Table 33), a similar pattern emerges. The parameter estimate between the non-hierarchical base (biscuit) is positive and the Bayesian confidence intervals and the frequentist measures indicate this is statistically so, indicating they are statistically significantly lower for both the hierarchical and non-hierarchical models. This means the Informational reinforcement within the higher utilitarian reinforcement group is influencing volume positively but less so than the biscuit category. For the beans category, there is disagreement between the two models, whereby the non-hierarchical model suggests the beans category also influences volume but in a lesser capacity to the base biscuit category and both Bayesian confidence intervals and frequentist t-statistics affirm this. However, the hierarchical model indicates there is no statistical difference between the beans and biscuit category as far as the value of this parameter is concerned, given the Bayesian confidence interval straddling zero and the non-significant t-statistic.

The constructed estimates and confidence intervals show variation between the hierarchical and non-hierarchical estimates though there is consistency in terms of the direction of the effect for biscuits, beans and fruit juice which are all statistically positive. The yellow fats category shows a negative overall effect for the non-hierarchical model and a zero effect for the hierarchical model.
This highlights the differences which can be deduced from choice of model structure.

7.3.2.4 Supermarket own brand x Informational reinforcement in the lower Utilitarian reinforcement group

SuperOwn x Informational x Ut1	Non-Hierarchical						Hierarchical					
	Estimate	Bayes CI	t-stat	sig	Constructed Est & CI		Estimate	Bayes CI	t-stat	sig	Constructed Est & CI	
Biscuits	0.007	0, 0.014	1.83	0.075	0.007	0, 0.014	-0.001	-0.008, 0.006	-0.37	0.373	-0.001	-0.008, 0.006
Fruit Juice Offset to Biscuits	-0.041	-0.055, -0.026	-5.59	0.000	-0.034	-0.039, -0.029	-0.032	-0.046, -0.019	-4.74	0.000	-0.034	-0.038, -0.029
Yellow Fat Offset to Biscuits	-0.064	-0.079, -0.049	-8.51	0.000	-0.057	-0.062, -0.052	-0.042	-0.056, -0.028	-5.83	0.000	-0.044	-0.048, -0.039
Baked Beans Offset to Biscuits	-0.120	-0.138, -0.102	-13.55	0.000	-0.113	-0.118, -0.107	-0.121	-0.138, -0.105	-14.38	0.000	-0.123	-0.128, -0.117

Table 34: Supermarket own brand x Informational reinforcement (lower Utilitarian group) - fixed effect

The base category is the biscuit category. From Table 34, the non-hierarchical model demonstrates some evidence this variable is statistically contributing to the model since the Bayesian confidence interval has the value zero at its lowest extremity of the interval. The t-statistic of t=1.83 is significant at p=0.075. This would imply the supermarket own brands within the lower utilitarian reinforcement group benefit volume-wise by having a higher informational reinforcement associated with their brands. Though it is worth noting this result is borderline given the lower estimate of the confidence interval is the zero value and the t-statistic is significant at 92.5% (not at 95%).

The hierarchical model, however, for the biscuit brands show the Bayesian confidence intervals straddling zero and a small and negative t-statistic of -0.37 (p=0.373) associated with the parameter. This would indicate the parameter is not statistically different from zero and hence the variable is having no effect on the volume per purchase within the category. Hence different interpretations of the variable are arrived at whether the hierarchical or non-hierarchical structure is observed.

The fruit juice category shows estimates which are negative for both model structures versus the biscuit base category and these are statistically robust given the Bayesian confidence intervals do not contain zero and the t-statistics are of a large magnitude and negative with both p<0.001. The Bayesian confidence intervals for the hierarchical and non-hierarchical models overlap, demonstrating the estimate is statistically similar. Also, the reconstructed confidence intervals of the estimate of the parameter are negative which implies the estimate of the variable is negative for this category. This implies supermarket own brands within the lower utilitarian reinforcement group see a negative relationship with their informational reinforcement. Hence consumers are seeking brands with lower informational reinforcement when shopping for supermarket own brands in the lower utilitarian reinforcement group. This may be due to a price orientated shopper where higher informational products are associated with higher price points.

Similar conclusions are drawn from inspection of the other two categories from Table 34. The yellow fats and the beans category show estimates which are negative for both the hierarchical and non-hierarchical model structures. These estimates show Bayesian confidence intervals which do not contain zero and also t-statistics which are statistically significant at $p<0.001$. In each category, the hierarchical and non-hierarchical confidence intervals overlap suggesting the estimates are of a similar magnitude. The reconstructed confidence intervals for both categories imply the estimates are negative and hence similar conclusions are reached as to the fruit juice category, whereby consumers are not seeking high informational reinforcement brands whilst shopping for supermarket own brands within the lower utilitarian reinforcement group of the yellow fat and beans category.

In conclusion, only the non-hierarchical model for the biscuit category would imply a positive relationship with this variable and the volume per purchase and this is a borderline relationship given the confidence interval extremity being zero and the t-statistic being relatively low. The equivalent hierarchical model suggests this effect is not statistically different to zero, whilst all other categories would imply the relationship of a negative nature to the volume per purchase. Therefore, in general, supermarket own brands within the lower utilitarian reinforcement group would benefit by appealing to a lower informational reinforcement strategy which may be due to associations between other informational reinforcement and higher prices which may not be what consumers within this lower equity group are seeking.

The hierarchical parameter estimates are showing signs of shrinkage versus the non-hierarchical with estimates ranging from (-0.121, -0.001), taking the maximum and minimum of the four categories, versus (-0.120, 0.007) for the hierarchical and non-hierarchical models respectively, in line with Rossi and Allenby (2003) who highlights the importance of model structure when determining model build.

7.3.2.5 Supermarket own brand x Informational reinforcement in the lower Utilitarian reinforcement group

SuperOwn x Informational x Ut2	Non-Hierarchical						Hierarchical					
	Estimate	Bayes CI	t-stat	sig	Constructed Est & CI		Estimate	Bayes CI	t-stat	sig	Constructed Est & CI	
Biscuits	-0.092	-0.103, -0.082	-17.75	0.000	-0.092	-0.103, -0.082	-0.080	-0.089, -0.07	-16.01	0.000	-0.080	-0.089, -0.07
Fruit Juice Offset to Biscuits	0.002	-0.029, 0.034	0.15	0.394	-0.090	-0.099, -0.081	0.037	0.007, 0.068	2.34	0.026	-0.043	-0.052, -0.033
Yellow Fat Offset to Biscuits	0.144	0.119, 0.169	11.30	0.000	0.052	0.044, 0.06	0.102	0.078, 0.125	8.34	0.000	0.022	0.014, 0.03
Baked Beans Offset to Biscuits	0.082	0.058, 0.105	6.74	0.000	-0.011	-0.018, -0.003	0.061	0.038, 0.083	5.32	0.000	-0.019	-0.026, -0.012

Table 35: Supermarket own brand x Informational reinforcement (higher Utilitarian group) - fixed effect

The supermarket own brands' informational reinforcement within the higher utilitarian group is an offset of the same variable within the lower utilitarian reinforcement group. The biscuit category is the base category (results shown in Table 35).

The hierarchical and non-hierarchical models of the biscuit category indicate a negative relationship between this variable and volume per purchase. The negative extremities of the Bayesian confidence intervals and the large magnitude and negative t-statistic illustrates this is statistically significant in model prediction. It implies consumers are negatively impacted by supermarket own brand's informational reinforcement within the higher utilitarian reinforcement group. The values for the hierarchical model for this category is similar to that of the non-hierarchical where a negative offset is present for the higher utilitarian reinforcement group versus the lower. Also, the constructed confidence intervals of both models overlap indicating similar levels between the two variables.

The non-hierarchical model offset for fruit juice (versus the biscuit category) is not different from zero given the Bayesian confidence interval is straddling zero and also the low t-statistic (p=0.394). However, the constructed confidence interval for the fruit juice category for the supermarket own brands informational reinforcement within the higher utilitarian reinforcement group is (-0.099, -0.081) which implies it is statistically lower than the lower utilitarian reinforcement group, in line with the non-hierarchical biscuit category. Again, consumers seem less interested in higher informational brands within this purchasing sector. The offset of the hierarchical model is higher than that of the offset of the hierarchical equivalent within the biscuit category, with the confidence interval not containing zero and the t-statistic being sufficiently large (p=0.026). This is a different conclusion from the non-hierarchical model. However, when the confidence intervals are constructed for the value of the lower utilitarian group, the confidence intervals are negative at both extremities indicating the offset to the lower utilitarian reinforcement group is negative (as seen with the non-hierarchical model).

The yellow fats category, however, shows a positive offset for both the hierarchical and non-hierarchical models versus the base biscuit category and this in turn is constructed to be a positive offset versus the lower utilitarian reinforcement group for the informational reinforcement of the supermarket own brands. This is in contrast to what has been observed in the biscuits and fruit juice category and shows a different dynamic to the category. Supermarket brands with higher informational reinforcement

From the beans category perspective, the non-hierarchical model shows the supermarket own brand effect within the higher Utilitarian group is statistically higher than the base biscuit category, though still an overall negative real terms effect versus the lower Utilitarian reinforcement group. The hierarchical model agrees in terms of direction, though the offset coefficient to the base category is smaller and hence the overall effect is still negative though with larger magnitude, i.e. a more negative effect. Hence, we see consumers less attracted to supermarket own brands with a higher Informational reinforcement in the higher Utilitarian reinforcement group.

7.3.2.6 Christmas effect x Informational reinforcement in the lower Utilitarian reinforcement group

Christmas x Ut 1	Non-Hierarchical						Hierarchical					
	Estimate	Bayes CI	t-stat	sig	Constructed Est & CI		Estimate	Bayes CI	t-stat	sig	Constructed Est & CI	
Biscuits	0.058	0.001, 0.116	1.96	0.058	0.058	0.001, 0.116	0.059	0.007, 0.113	2.19	0.036	0.059	0.007, 0.113
Fruit Juice Offset to Biscuits	-0.054	-0.144, 0.034	-1.18	0.198	0.004	-0.055, 0.064	-0.045	-0.125, 0.036	-1.10	0.217	0.014	-0.039, 0.067
Yellow Fat Offset to Biscuits	-0.122	-0.202, -0.04	-2.95	0.005	-0.064	-0.112, -0.016	-0.126	-0.199, -0.054	-3.38	0.001	-0.068	-0.112, -0.023
Baked Beans Offset to Biscuits	-0.087	-0.195, 0.017	-1.59	0.112	-0.029	-0.088, 0.03	-0.067	-0.166, 0.03	-1.37	0.157	-0.008	-0.062, 0.045

Table 36: Christmas effect x Informational reinforcement (lower Utilitarian group) – fixed effect

From Table 36, the point estimate for the base biscuit category for both models is similar at 0.59 (hierarchical) and 0.58 (non-hierarchical) models. The confidence intervals of both models do not contain zero and both intervals overlap suggesting the estimate is similar for both models. The t-statistics are both significant at $p<0.059$. This suggests that volume per purchase is higher within the Christmas week than another average week.

The Bayesian posterior confidence intervals for the fruit juice category straddle zero for the hierarchical and non-hierarchical models and the t-statistics are statistically non-significant from a frequentist perspective ($p=0.22$ and $p=0.19$ for the hierarchical and non-hierarchical models respectively). This implies there is no difference, statistically, between the estimates of the fruit juice variant of the parameter and the (base) biscuit parameter. However, the constructed parameter estimate in both cases is low (0.014 and 0.004 respectively) and the constructed confidence interval straddles zero in both cases, implying the parameter is not different from zero (at least statistically). This demonstrates the importance of interpretation of the parameter since the initial interpretation would be no difference from the biscuit parameter (which is deemed to be positive); however, given the wider confidence of the fruit juice offset parameter, this has had the effect of widening the confidence interval of the constructed parameter to suggest a high posterior probability the parameter is zero. These differences make it challenging to inform a lay audience management as to what assumptions should be made when modelling scenarios based on this complex approach.

The yellow fat offset for the hierarchical model is -0.126 and for the non-hierarchical model is -0.122, hence similar results. The Bayesian confidence intervals imply very low probability the parameter is zero in each case (supported by the high t-statistics, both $p<0.006$). The extremities of the confidence intervals of the constructed parameter are negative for both models and their values are very similar suggesting similarity between the hierarchical and non-hierarchical models. This implies the offset for the yellow fats category for the parameter is statistically lower than that of the base biscuit category and furthermore the estimate within the model is negative. Therefore, within the yellow fats category, the Christmas week has a negative impact on volume per purchase compared to other average weeks within the year. The beans category offset for both the hierarchical and non-hierarchical models contains the value zero indicating a high probability the parameter is zero. Also, the t-statistics are small giving significance levels of $p=0.157$ and $p=0.112$ respectively. Therefore, the beans offset is not statistically different from that of the base biscuit category. The constructed confidence intervals however, both for the hierarchical and non-hierarchical models straddle zero which indicate the parameter is not statistically different from zero. Hence the same pattern is seen as with the fruit juice category, where the base category confidence intervals are positive, the offset is non-statistically different from zero, however the constructed confidence intervals suggest the parameter is zero (different from that of the base biscuit category). As with the fruit juice category, this is due to the comparably larger offset confidence intervals (compared to the biscuit category) which has the effect of widening the constricted confidence interval to straddle zero.

This parameter shows much consistency between the two model functional forms. Both the hierarchical and non-hierarchical models have led to the same conclusion in terms of whether the offset and the constructed confidence intervals contain the value zero or not. Also, where the constructed intervals point to a non-zero relationship between the parameter and the volume per purchase, both models agree on the direction of this relationship and also the magnitude (i.e. overlap in the confidence intervals).

However, this parameter has also demonstrated instances whereby the offset inference and the constructed estimate inference can bring differing conclusions.

7.3.2.6 Christmas effect x Informational reinforcement in the lower Utilitarian reinforcement group

Christmas x Ut 2	Non-Hierarchical						Hierarchical					
	Estimate	Bayes CI	t-stat	sig	Constructed Est & CI		Estimate	Bayes CI	t-stat	sig	Constructed Est & CI	
Biscuits	0.008	-0.076, 0.092	0.17	0.393	0.008	-0.076, 0.092	-0.018	-0.097, 0.06	-0.43	0.363	-0.018	-0.097, 0.06
Fruit Juice Offset to Biscuits	0.069	-0.136, 0.28	0.65	0.323	0.077	0.011, 0.143	0.097	-0.091, 0.283	1.03	0.235	0.079	0.02, 0.138
Yellow Fat Offset to Biscuits	0.064	-0.093, 0.218	0.81	0.288	0.072	0.016, 0.128	0.050	-0.095, 0.197	0.67	0.318	0.033	-0.019, 0.084
Baked Beans Offset to Biscuits	0.170	-0.033, 0.373	1.63	0.106	0.177	0.112, 0.243	0.176	-0.007, 0.37	1.86	0.071	0.158	0.099, 0.218

Table 37: Christmas effect x Informational reinforcement (lower Utilitarian group) – fixed effect

The base biscuit value for the Christmas effect week within the higher utilitarian reinforcement group (shown in Table 37) is not statistically different from zero given the Bayesian confidence interval and the small t-statistics.

With regards to the beans category, compared to that of the biscuit category, both hierarchical and non-hierarchical have a positive point estimate for this parameter and both are similar in magnitude (0.176 and 0.170 in the usual order). However, the confidence intervals of the parameter are relatively wide for both models and hence, despite the large magnitude of the point estimate, from a Bayesian inference perspective, there is high probability this estimate is zero for both the hierarchical and non-hierarchical models. The frequentist t-statistics concur with small t-statistics resulting in non-significant t-statistics ($p=0.071$ and 0.106 respectively). The hierarchical model is significant at $p<0.072$ and hence there is evidence this offset effect is statistically significant and indeed the Bayesian confidence interval is close to zero. If this parameter is accepted as being a significant contributor to the model, then the hierarchical and non-hierarchical models will differ in conclusion in terms of offset significance despite the point estimates being very similar. The constructed confidence intervals for the parameter are both positive and they also overlap suggesting agreement between the two models. This suggests the overall effect of Christmas week within the lower Utilitarian reinforcement group for the beans category has a negative effect on volume per purchase.

7.4 Comparison of Pooled and Fixed Effect Models

Fig 103 forms a graphical representation of the four combined category models (pooled and fixed effect). The grey bars show the fixed effects model (i.e. those where the coefficients are allowed to vary for each category) where the solid bars are the hierarchical model estimates and the lined bars the non-hierarchical estimates. The blue bars towards the right of the chart are the pooled estimates for both the non-hierarchical and hierarchical models. Visualising the charts, it can be seen the pooled estimates are an average of sorts of the fixed effects models. This offers support for the earlier argument of a generalisation of the pooled coefficient over the four categories and while this may be useful as a strategic tool to management to assess the nature of a cross-category effect, the potential danger is the averaging of the coefficient may disguise any category specific deviations which the fixed effect model would uncover.

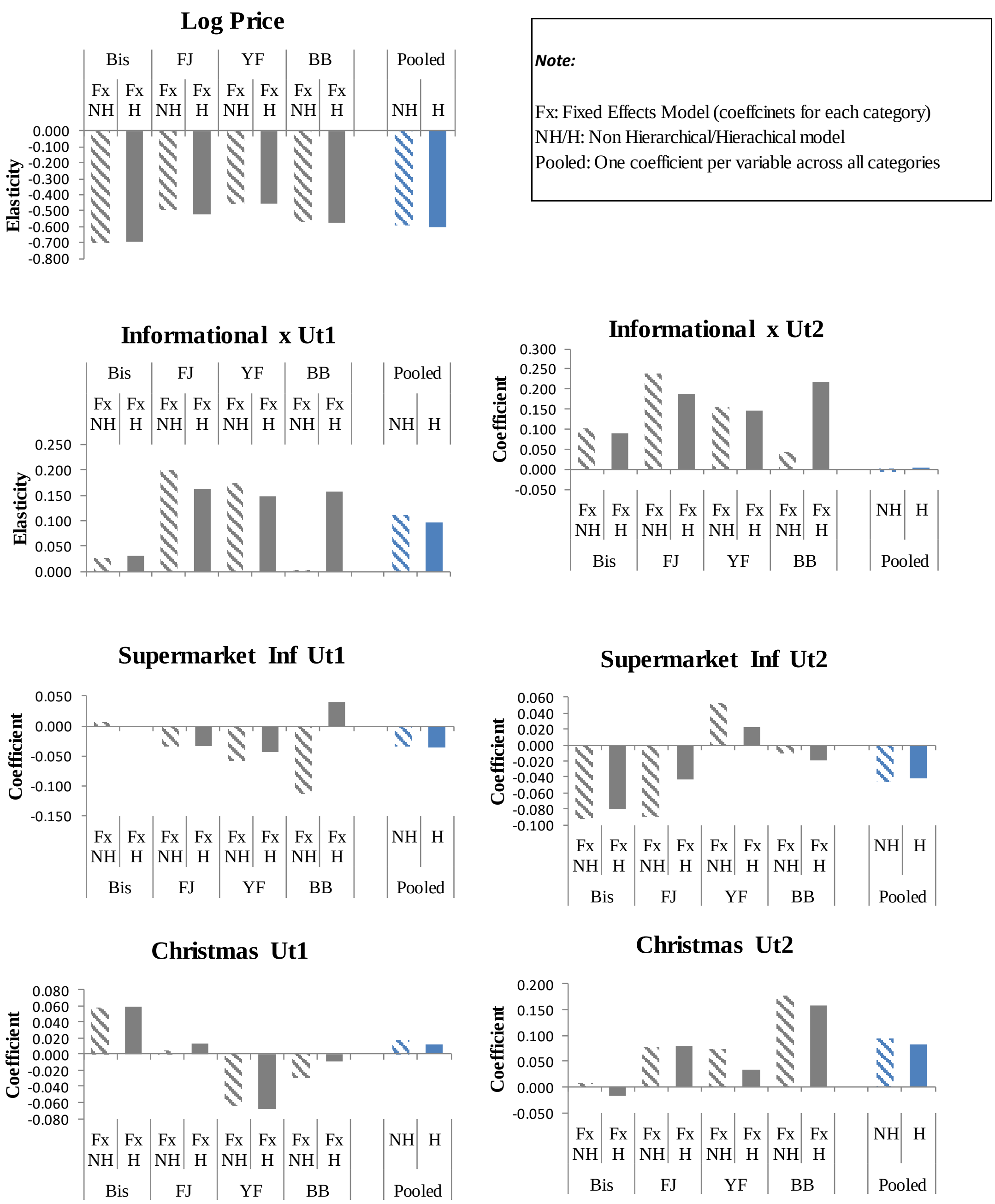

Figure 103: Comparison of pooled and fixed effect coefficients column chart

The actual statistics which feed Fig 103 are sown below in Table 38.

	Bis		FJ		YF		BB		Pooled	
	Fx NH	Fx H	Fx NH	Fx H	Fx NH	Fx H	Fx NH	Fx H	NH	H
Log Price	-0.701	-0.696	-0.493	-0.524	-0.456	-0.458	-0.570	-0.576	-0.591	-0.602
Informational x Ut1	0.027	0.032	0.200	0.161	0.173	0.148	0.002	0.157	0.111	0.097
Informational x Ut2	0.100	0.089	0.239	0.186	0.155	0.146	0.043	0.217	-0.006	0.001
Supermarket Inf Ut1	0.007	-0.001	-0.034	-0.034	-0.057	-0.044	-0.113	0.039	-0.033	-0.036
Supermarket Inf Ut2	-0.092	-0.080	-0.090	-0.043	0.052	0.022	-0.011	-0.019	-0.046	-0.042
Christmas Ut1	0.058	0.059	0.004	0.014	-0.064	-0.068	-0.029	-0.009	0.018	0.012
Christmas Ut2	0.008	-0.018	0.077	0.079	0.072	0.033	0.177	0.158	0.093	0.082

Table 38: Comparison of pooled and fixed effect coefficients

7.4.1 Comparing Model Performance

Within this section, the categories have been run in one model rather than as separate models. The models have varied by some being pooled and some with fixed effects. Also, both have been run as hierarchical and non-hierarchical models. The results show all model variations show a good representation of the underlying data, however consideration of which model may better represent the combined categories model is now discussed.

There are arguably two considerations to assess and balance. The first is the underlying statistical model diagnostics in assessing how well the model represents the underlying data. It has been argued that all model functional forms represent the data however there is a hierarchy in terms of model favourability, purely from a diagnostic perspective. Table 39 is a summary of the model diagnostics of the four models, pooled and fixed effect with hierarchical and non-hierarchical variants within both. There is a consistency between the Bayesian and frequentist diagnostics whereby models with better Bayesian diagnostics also have better frequentist diagnostics. This fact does underline Efron, (2005) view that a pluralism view of statistics whereby Bayesian and frequentist methods are used in complement to each other is a preferred.

Employing this logic, Table 39 is appended with an additional statistic relating to the author's ranking of the models, based on the model diagnostics alone. The hierarchical models rank first and second with the fixed effects model being preferred to the pooled model. The same pattern is seen for the non-hierarchical models. Hence the hierarchical structure is more important than the pooled/fixed effects structure of the model (at least statistically). An interesting finding is the similarity of the diagnostic measures across the four models. Despite very different model structures, all four models are explaining the data with similar diagnostic (e.g. R-squared (adj) ranging between 69.447% and 72.858%, DIC between

179,169 and 154,087) hence, statistically all models could be considered as good representations and almost equally good representations of the underlying data.

	Pooled Non Hierarchical	Pooled Hierarchical	Fixed Effect Non Hierarchical	Fixed Effect Hierarchical
R-Squared (adj)	69.447%	72.202%	70.435%	72.858%
Mean Deviance	179,127	154,772	176,177	152,502
Penalty	42	1,564	63	1,584
DIC	179,169	156,336	176,240	154,087
MAPE	6.636%	6.207%	6.534%	6.126%
Variance (between purchases)	0.240	0.198	0.235	0.195
Variance (between housholds)		0.047		0.045
Variance (between household) t-stat (sig)		58.358(0)		57.881(0)
Variance Partition Coeficient		19.150%		18.766%
Rank of prefered model	4	2	3	1

Table 39: Model diagnostics comparison across all combined models

A benefit for improving the model fit (statistically) would be to run it as a hierarchical structure rather than a non-hierarchical. Also, there is consistency in that the fixed effect model is superior to the pooled model. The model diagnostics suggest the pooled hierarchical model is a (statistically) better representation of the data than a fixed effect non-hierarchical model. However, consider the data which has been analysed within this study. The data represents FMCG non-luxury purchases and despite being physically very different products, in terms of households purchasing, all could be considered to be very similar in nature. That is, a product which can be purchased with relatively little disposable income, high distribution levels in terms of weighted commodity value, a relatively fixed purchase cycle and resident in a competitive environment. It may be the case, therefore, that the fact the pooled model is better than the fixed effect model is the coefficients of the products have little variation.

It may be that more and maybe quite different categories should be analysed in this fashion before making the generalisation of this kind.

The second consideration relates to the practicalities of the models. The data relate to four FMCG categories which are distributed through large retailers (in the UK) and require individual brand strategies. The nature of the pooled models means the coefficients are generalized across the categories which may hide cross category differences between the focal parameters, as small differences from the average may influence cost structure, promotional points or marketing deployment. Also, from a practitioner's perspective, retailer discussion may demand model estimates to be category based rather than a pooled average which may favour a fixed effects model. The nature of the similarity of estimates of these four categories has endeared them to the structure of a pooled model. Given the range of FMCG categories available within the UK marketplace, it would be unlikely the relative homogeneity would span all categories and hence the pooled nature of applying one estimate to a parameter across all categories would seem illogical. Also, there has been opinion in the literature which calls for a fixed effects approach over a pooled approach (Cameron and Trivedi, 2010). Therefore, from a practical perspective, it is argued a fixed effects model would be preferred to a pooled model, even though this increases the complexity of the model.

For each model the hierarchical functional form has seen an increase in model performance, though the increase is not as significant as was observed with the separate category models. From a behavioural perspective, it is also more intuitive that purchases within household are not assumed to be independent for reasons discussed within the methodology chapter. The combined category model has taken this a stage further since the household spans category and hence the notion of independence of category has been removed. Whereas within the separate category models, each category was treated as completely independent by the nature of the different models. A comparison of how this impacts the parameter estimates and hence how a practitioner would act based on the results is discussed next.

7.5 Combintion vs. separate model comparison

The data have been modelled as four separate categories and as one large model. Given the hierarachical models in both these separate and combined models are prefered, the study discusses the differences between the two models. The fixed effects model is considered for the combined model given better model diagnostics, prefered pragmatic results in terms of category specific coefficients and also the ability to compare category by category rather than each separate caregory model coefficients with the one pooled coefficient.

Fig 104 shows a graphical representation of the coefficients for the focal parameters. The fixed effect coefficients have been constructed to reflect the equivalent to the separate model, for example the informational reinforcement for the fruit juice higher utilitarian group is the combination of the base category parameter and the fruit juice offset specific to this parameter. This ensures comparability between the separate and combined model parameter estimates.

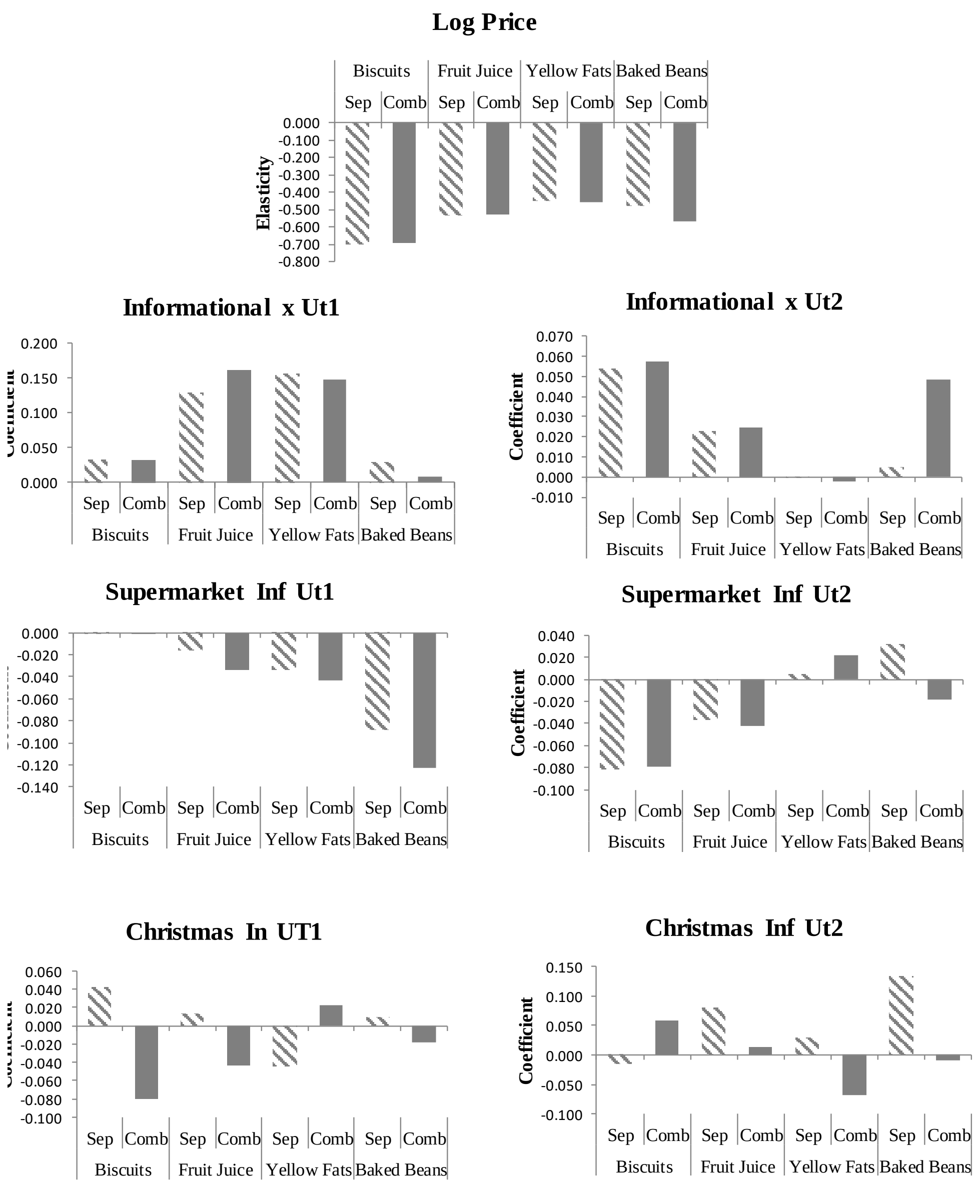

Figure 104: Coefficients of combined and separate models column chart

Table 40 shows the actual values that constitute the chart, along with the Bayesian posterior confidence interval which gives the ability to see if the coefficients are significantly different to zero. The blue cells highlight the coefficients which are statistically significantly negative (upper and lower Bayesian confidence interval is negative). The red which are statistically significantly positive (upper and lower Bayesian confidence interval is positive) and the white cells which are not statistically different from zero (confidence interval straddles zero). This allows a pictorial view of the similarities between the separate and combined category models from a directional perspective (not magnitude). There are 4 categories and 7 metrics and hence 28 comparisons across the separate and combined categories. Of the 28, 25 (89.3%) are in agreement directionally between the separate and combined models and 3 in disagreement (10.7%).

Price is the most consistent when comparing the separate and the combined models, in agreement for all categories. The magnitude is also similar and the confidence intervals of the separate and combined models overlap in all categories except that of biscuits. Given the nature of the categories (commodity FMCG), it would be surprising if there was a non-negative relationship present.

Informational Reinforcement in the both the lower and higher utilitarian groups is consistent between the separate and combined categories and hence the BPM core variables are important in understanding the influence on behaviour

.

There is some disagreement within the supermarket set of variables and also the Christmas set of variables when comparing them between the separate and combined category models though by in large there is agreement directionally.

The supermarket own brand informational reinforcement generally is negative across all categories and all models, though not completely consistent across all.

The Christmas effect is generally a zero effect with the exception of some models. Hence looking across both sets of models there seems to be a general consensus of the following

Price: Negative effect
Informational/Utilitarian reinforcement: Generally positive effect
Supermarket own effect on the BPM variables: Generally negative effect
Christmas week effect on the BPM variables: Generally zero effect

	Biscuits				Fruit Juice			
	Sep	Sep CI	Comb	Comb CI	Sep	Sep CI	Comb	Comb CI
Log Price	-0.702	-0.71, -0.695	-0.696	-0.703, -0.688	-0.531	-0.549, -0.513	-0.524	-0.53, -0.519
Informational x Ut1	0.033	0.026, 0.039	0.032	0.025, 0.038	0.128	0.111, 0.145	0.161	0.156, 0.166
Informational x Ut2	0.054	0.046, 0.062	0.058	0.05, 0.066	0.023	-0.006, 0.052	0.025	0.017, 0.032
Supermarket Inf Ut1	-0.001	-0.008, 0.005	-0.001	-0.008, 0.006	-0.017	-0.03, -0.003	-0.034	-0.038, -0.029
Supermarket Inf Ut2	-0.081	-0.09, -0.071	-0.080	-0.089, -0.07	-0.037	-0.07, -0.003	-0.043	-0.052, -0.033
Christmas	0.043	-0.009, 0.094	0.059	0.007, 0.113	0.014	-0.045, 0.074	0.014	-0.039, 0.067
Christmas Ut1	-0.015	-0.094, 0.067	-0.018	-0.097, 0.06	0.080	-0.09, 0.258	0.079	-0.097, 0.06

	Yellow Fats				Baked Beans			
	Sep	Sep CI	Comb	Comb CI	Sep	Sep CI	Comb	Comb CI
Log Price	-0.448	-0.462, -0.435	-0.458	-0.463, -0.452	-0.476	-0.499, -0.453	-0.570	-0.578, -0.563
Informational x Ut1	0.157	0.146, 0.168	0.148	0.144, 0.153	0.029	0.014, 0.044	0.008	0.003, 0.013
Informational x Ut2	-0.001	-0.009, 0.007	-0.002	-0.007, 0.003	0.005	-0.008, 0.017	0.049	0.044, 0.054
Supermarket Inf Ut1	-0.033	-0.045, -0.021	-0.044	-0.048, -0.039	-0.088	-0.107, -0.07	-0.123	-0.128, -0.117
Supermarket Inf Ut2	0.004	-0.017, 0.025	0.022	0.014, 0.03	0.031	0.009, 0.054	-0.019	-0.026, -0.012
Christmas	-0.044	-0.085, -0.002	-0.068	-0.112, -0.023	0.010	-0.063, 0.08	-0.008	-0.062, 0.045
Christmas Ut1	0.030	-0.069, 0.13	0.033	-0.097, 0.06	0.133	-0.011, 0.277	0.158	-0.097, 0.06

Table 40: Combined and separate model coefficients

Despite the consistency (at least directionally) between the separate and combined models, it is again worth noting the differences between the models and how the choice of how models are structured and run can influence the end results; hence the importance of model structure.

7.6 Summary of the Models.

Data relating to four FMCG categories have been analysed within this study and following initial analysis several models have been built. In the first instance, separate models were built for each category, hence assuming independence between the categories and allowing the behaviour across them to vary. Models were built using a hierarchical and a non-hierarchical structure. The hierarchical models were divided into models with vague prior distributions and models with informative prior distributions.

Diagnostically, the hierarchical models were a better representation of the underlying data than the non-hierarchical models. From a theoretical perspective, the hierarchical structure seemed more relevant to the nature of the data given the purchase within household hierarchy of the data.

The informative prior distributions, in some cases, were constructed using a high level of precision, much of which was driven by the large sample of the data. Given the fact this type of modelling is newly applied within the BPM, a vague prior distribution was favoured for this study, however, these prior distributions could form a basis of more informative studies ongoing (O’ Hagan, 1994). Also, the categories within the study are not intended to generalise to a wider FMCG categories given the vast diversity of the products associated with the FMCG overarching category.

The weakness of the hierarchical model, however, is the lack of ability to accurately forecast future volume projections. This is due to the fact the panel id which represents the household within the panel data is assigned a coefficient and it would be difficult to know which panel id coefficient should be applied to the new projected data point. It is argued an average across all may be a suitable estimate. The non-hierarchical does not suffer from this issue though the model diagnostics were weaker in all categories. It was argued the hierarchical model with vague priors be the recommendation specifically for this study.

The study progressed to build a combined study across all four categories. This reflected the fact that the assumption of independence across category could be incorrect assumption, since the majority of households (86%) purchased from more than one of the four product categories. The model was built to reflect a hierarchical and a non-hierarchical model. Given the complexity of deriving initial values for the numerous offsets, only a vague prior distribution hierarchical model was constructed, which highlights both the difficulty of the added complexity of Bayesian models, however, also highlights the power of the vague prior in dealing with instances where either the prior distribution is too complex to calculate or where prior knowledge is not available. The results favoured the hierarchical model (at least diagnostically) over the non-hierarchical model.

For the combined category model, a pooled model and an offset model were considered. There was little difference in the diagnostic results of the two models. However, it was argued this was down to the homogeneity of the underlying parameters of the four categories modelled. Other categories with differing elasticity of behavioural psychological parameter estimates would be “averaged” using the pooled model and hence more categories would need to be modelled in order to understand the potential range of these parameter values.

From a practitioner’s perspective, it would also seem more useful to be able to negotiate with retailers and marketing agencies using category specific results rather than cross-category results.

Finally, the hierarchical fixed effect model with vague prior was compared to the hierarchal models (all with vague priors) for the separate categories. Directionally the parameters were consistant with (89.3% of parameters directional agreement). However, this does mean some differ in direction and most in magnitude. The diagnostics of both models were comparable and hence the decision to favour the combined category model was made based on the basis the independence of households purchasing across-category should not be ignored.

Chapter 8: Discussion and Future Research

8.1 Discussion of the research questions

8.1.1 Discussion of the research questions relating to separate categories

8.1.1.1 RQ1

The first question was to test the economic behaviour price variable and how the differing model structures may affect the interpretation of the parameter. For all three models, the price elasticity (in Fig 105) is similar and also similar to other studies involving the BPM with different functional forms (Oliveira-Castro *et al.*, 2006; Chang, 2007). The inclusion of a more complex model or Bayesian estimation has not changed the fundamental understanding of the price elasticity measure. This underlines the benefit of the BPM which allows economic behaviour to be included alongside the psychology variables of the BPM without collinearity impacting the values of the price elasticity variable.

Therefore, RQ1 is deemed to show the elasticity of demand is apparent within the more complex model structure; is unaffected by the model hierarchical structure or and prior distribution; and the Bayesian inference is returning estimates comparable with past studies.

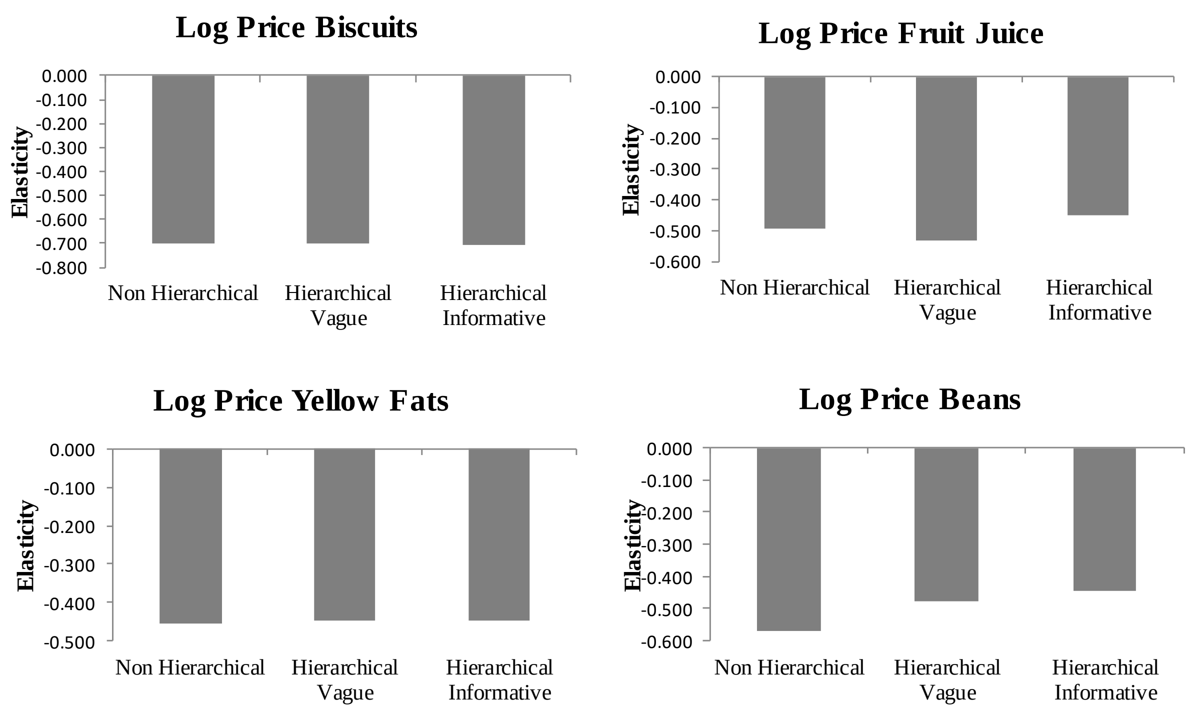

Figure 105: Price elasticity coefficients column chart

Fig 106 shows a colour coded summary of the parameter signs. Interpretation of the code is as follows. The first symbol is the sign of the variable (+/-). The second symbol (^) indicates if the Bayesian confidence intervals of the posterior estimate do not include the value zero, this symbol is absent if they do. The next symbol (**/*) indicates the parameters is statistically significant at 5% level (**) or 10% level (*) or absent if not. A red colour indicates statistically significant and positive, blue indicates statistically significant and negative and lack of colouring indicates not statistically significant.

For the price elasticity of demand variable, all categories and all model forms are negative and statistically significant from a Bayesian and frequentist perspective.

	Non Hierarchical	Hierarchical Vague	Hierarchical Informative
Biscuits	- ^**	- ^**	- ^**
Fruit Juice	- ^**	- ^**	- ^**
Yellow Fats	- ^**	- ^**	- ^**
Beans	- ^**	- ^**	- ^**

Figure 106: Price comparison

8.1.1.2 RQ2

The second research question focussed on the Informational and Utilitarian reinforcement variables of the BPM. These have been shown to be statistically significant in determining consumer behaviour in past studies (e.g. Oliveria-Castro *et al.*, 2008). Considering the more complex model within this study where the BPM variables are being used in context of other interaction variables (i.e. supermarket own brand and the Christmas seasonality), there is a requirement to assess the statistical relevance of these variables within the more elaborate model framework. The results are discussed in depth in the analysis body of the study. The graphical overview in Fig 107 below summarises the relationship between purchases and the informational reinforcement within the lower utilitarian reinforcement group. It shows a positive and statistically significant result for all categories within the hierarchical vague models. This is true for the hierarchical informative and non-hierarchical models, with the exception of beans. The beans category highlights the importance of choice of model functional form and also prior distribution since the three models result in all possible forms of the relationship between the variables, i.e. neutral, positive or negative.

	Non Hierarchical	Hierarchical Vague	Hierarchical Informative
Biscuits	+ ^**	+ ^**	+ ^**
Fruit Juice	+ ^**	+ ^**	+ ^**
Yellow Fats	+ ^**	+ ^**	+ ^**
Beans	+	+ ^**	- ^**

Figure 107: Informational reinforcement (lower Utilitarian group) comparison

Similarly, an overview of the Informational reinforcement within the higher (second) utilitarian group is shown below in figure 108. This is an offset parameter to the Informational variable within the lower Utilitarian reinforcement level. The biscuit category has aligned estimates in terms of sign with all three models depicting a positive and a statistically relevant relationship. Other categories show much variation in how the parameter is portrayed, ranging from positive, neutral and negative relationships, depending on the model selected. The non-hierarchical model tends to show a positive relationship, the hierarchical vague tends to be neutral relationship and the hierarchical Informative a negative relationship, again highlighting the importance of the model structure (hierarchical or non-hierarchical) and also the nature of the prior (vague or Informative).

	Non Hierarchical	Hierarchical Vague	Hierarchical Informative
Biscuits	+ ^**	+ ^**	+ ^**
Fruit Juice	+ ^*	+	- ^**
Yellow Fats	- ^**	-	- ^**
Beans	+ ^**	+	- ^**

Figure 108: Informational reinforcement (higher Utilitarian group)

It can be concluded the BPM variables are contributing to the consumer behaviour across all four categories, albeit the interpretation can vary depending on the assumptions made about the independence of purchases within household (i.e. the hierarchical nature of the data) and the nature of the prior information applied to the data (i.e. informed versus vague).

8.1.1.3 RQ3

This research area focussed on the inclusion of a supermarket own brand interaction term with both the informational and utilitarian reinforcement elements of the BPM. The interaction was constructed using a base and an offset variable (shown in Fig 109). The base represented the impact of the supermarket own brand effect on the informational reinforcement variable. The offset represented the impact of the Informational reinforcement for the higher Utilitarian reinforcement group (i.e. an offset to the base variable).

SuperOwn x Informational

	Non Hierarchical	Hierarchical Vague	Hierarchical Informative
Biscuits	+ ^*	-	+ ^**
Fruit Juice	- ^**	- ^*	+ ^**
Yellow Fats	- ^**	- ^**	- ^**
Beans	- ^**	- ^**	-

SuperOwn x Informational GP2

	Non Hierarchical	Hierarchical Vague	Hierarchical Informative
Biscuits	- ^**	- ^**	- ^**
Fruit Juice	- ^**	- ^*	- ^**
Yellow Fats	+ ^**	+	-
Beans	-	+ ^*	+

Figure 109: Supermarket own brand x Informational reinforcement (lower and higher Utilitarian group)

Considering both the base variable (i.e. Informational within the lower Utilitarian reinforcement) and the offset (i.e. Informational within the higher Utilitarian reinforcement group) there is a difference separating the biscuit category from the other three. The biscuit category shows a positive relationship between the informational reinforcement for the supermarket own brand, within the lower utilitarian reinforcement group, while most models of the other category shows a negative relationship.

This suggests consumers of the biscuit category are actively seeking a supermarket own brand offering whilst shopping amongst the lower utilitarian brands and are being negatively influenced by the offering for higher utilitarian group brands. Hence consumers within the biscuit category are looking for higher informational branded treats when seeking higher utilitarian reinforcement biscuits. This would represent an opportunity for managers to entice an up trade to a higher priced offering for higher utilitarian reinforcement group biscuits.

For the other three categories, there is strong evidence the inverse is true. Hence when consumers purchase within the lower utilitarian group of supermarket own products, there is little appetite for higher informational reinforcement products, hence it could be argued the consumer is seeking basic informational reinforcement within this lower utilitarian and supermarket own product sector.

Outside of the biscuit category and for the higher utilitarian group of the supermarket own brands, there is a mixed set of results. Amongst the statistically superior hierarchical models the general trend is a neutral effect between the informational reinforcement and purchases. Hence the indication to management here within the fruit juice, yellow fats and beans category would be less emphasis on supermarket own products with higher informational reinforcement status.

This shows the power of the BPM as it allows marketers to understand the relationship between categories and how differing informational and utilitarian reinforcement can impact purchase patterns within the supermarket own group of brands. This allows for better category wide product offerings to meet consumers' psychological needs.

8.1.1.4 RQ4

The second interaction variable focusses on the seasonal Christmas week, having uncovered a significant drop in total volume for that period within the category analysis section across all categories within the study. As with the supermarket own brand, the Christmas variable is divided into a base (the interaction with the informational reinforcement in the lower utilitarian group) and an offset (the additional impact of the informational reinforcement within the higher utilitarian group for the Christmas week). Fig 110 shows the Christmas week interaction charts for the informational reinforcement within both the lower and higher utilitarian informational reinforcement groups.

Chrsitmas

	Non Hierarchical	Hierarchical Vague	Hierarchical Informative
Biscuits	+ ^	+	+
Fruit Juice	+	+	+
Yellow Fats	- ^*	- ^*	- ^**
Beans	-	+	+

Chrsitmas Ut Gp2

	Non Hierarchical	Hierarchical Vague	Hierarchical Informative
Biscuits	+	-	+
Fruit Juice	+	+	+
Yellow Fats	+	+	-
Beans	+	+	+ ^

Figure 110: Christmas effect x Informational reinforcement (lower and higher Utilitarian group)

The only consistent indication of a relationship is a negative one for the informational reinforcement in the yellow fats category within the lower utilitarian reinforcement group. Hence despite the volume drop at Christmas, for those who are shopping, their psychological pattern does not change, apart from the yellow fats category, where less importance is given to products with a higher informational reinforcement within the lower utilitarian reinforcement group. The other three categories show no effect in terms of psychological behaviour. Hence the Christmas period does not have a significant effect on individual consumer behaviour as far as the BPM variables are concerned. It is likely the decrease in volume can be attributed to fewer shoppers (buying in the same manner) and also the fewer shopping days within the period.

The implications to management is not to focus on these categories during the Christmas period as it is unlikely to affect consumers' purchasing patterns from a consumer psychology standpoint. The exception would be the yellow fats where less shelf space may be devoted to higher informational reinforcement products which reside within the lower utilitarian reinforcement group. However, given the relatively short period and the reduced volume being sold in this period, it may not be worth the resource expenditure to rearrange shelf space etc. solely for this week.

8.1.1.5 RQ5

RQ5 focussed on the diagnostic difference a hierarchical structure would bring to the model and how the interpretation of the model may differ. In order to address this, the study compared the non-hierarchical model to the hierarchical model with vague priors, as this isolates the difference attributed by the hierarchical structure alone. The structure is compared across the four categories. The model diagnostics across all four categories statistically improve with the hierarchical structure, from both a Bayesian and frequentist stance. This is due to the variance within household term being significantly different from zero and hence accounting for some of the variability resulting from the intra-household structure of the data. This is happening as the assumption of independence within household is refuted.

The hierarchical structure tested within this study is of the simplest form and relates to the intercepts of the households, i.e. common households have a common intercept term. This means there is a common intercept term for the category and offsets to this intercept relating to each of the number of unique households (h) within the specific data set. The random effects of this hierarchical structure means that a variance term is applied to this offset parameter and hence just 1 degree of freedom is required rather than $h-1$ degrees of freedom under a fixed effects hierarchical structure. The implication for this study may not be great given the large number of n observations though needs to be considered for future studies. Also, the Bayesian MCMC estimation requires the Gibbs sampler to estimate each parameter in a systematic manner which requires each parameter to be evaluated whilst keeping the others static and hence increased degrees of freedom would have a substantial impact on the resources required to estimate these parameters.

The notion of extending the hierarchical structure to include random slopes (as well as random intercepts) will be discussed later in this chapter when future considerations and limitations are discussed.

It can be concluded the hierarchical model is addressing the structure of the data in a statistically superior way, in terms of model diagnostics. The diagnostics of all models, however, are statistically sound and the most appropriate model is not always the one which offers a better statistical fit. A researcher may accept a slightly reduced set of model diagnostics for a model where coefficients are more intuitive or even where coefficients are not contradictive of what is regarded as fact.

The parameter inferences for both models are statistically relevant, which contribute to the models being diagnostically sound. The estimates are often different in magnitude and sometimes directionally different when comparing the hierarchical to the non-hierarchical models which affects interpretation of the results of course. This underlines the importance of model structural choice when designing and building models. It also poses a potential issue when dealing with managers with less experience of working with multiple statistical models and this will be addressed a little later in the chapter.

8.1.1.6 RQ6

The next research area focusses on the impact of an informative prior distribution versus a vague prior distribution. For this comparison, the two hierarchical models were compared since the hierarchical structure provides a better representation of the data for every category analysed within this study.

The nature of the prior distribution is very much within the control of the researcher and experiments can be set up with differing prior distributions. The area selected for this study was to follow a calibrated Bayes perspective, whereby the prior distribution was taken from the parameter estimate of regression based models. Whether this is a good choice of prior distribution and whether more relevant choices exist may come down to both the philosophy of the researcher or through experience of running similar studies. Where the prior distribution agrees with the likelihood, results have very similar parameter estimates for both the vague and informative models. This can be seen with the price elasticity coefficients which are statistically the same for both models. A contrast to this is the Informational reinforcement within the higher utilitarian reinforcement group, where for three categories (fruit juice, yellow fats and beans) there is zero effect for the vague models and negative effect for the informative models. This is due to a negative mean distribution with high precision being applied to the prior distribution of these three categories. This shows the influence the prior distribution can have on the estimate of the parameter, hence the importance the prior distribution plays.

The prior distribution could take an interesting role if models are updated on a rolling basis, where historical information about the parameter estimates can be used as a basis of ongoing prior distributions of future models. Instead of using one parameter regression models in a calibrated manner, these could be taken from historical studies where the prior itself benefits from an ongoing analysis of past results. It may also be that seasonality plays a role in whether priors should have larger or smaller precisions

For these four categories, the diagnostics are very similar for both the hierarchical vague and hierarchical informative models. There is an added complexity to the informative models given the offset approach, whereby the offset is reliant on the base estimate and changes in the base through the addition of the likelihood could mean the informative prior for the offset may not be relevant. For example, consider a situation whereby the base value of the variable has a prior distribution with a positive mean and the offset of the variable has a prior distribution with a negative mean. If, with the addition of the likelihood, the base estimate turned out to be negative, then this could potentially mean the negative nature of the offset prior distribution is no longer relevant. Hence the setting of informative priors to complex models can become tricky. When the study moved on to a combined category model, there was implicitly a further complication as each model is itself an offset to the biscuit category hence the number of moving parts increases, exacerbating the issue outlined above.

Therefore, given the added complexity of the informed model and the very little difference diagnostically, the combined category model focussed on the differences between the hierarchical and non-hierarchical, both with vague prior distributions.

8.1.2 Discussion of the research questions (combined categories)

The next stage of the study was to combine the categories into one model. This was done in four ways utilising a pooled model and a fixed effects structured model. The consumer panel is structured whereby the same households are included for all four categories and hence it is possible that a household could have purchased from more than one category in the fifty two week period. Hence the hierarchical concept used in the separate category analysis, whereby the assumption of independence between households was removed, can be extended to the combined model. This extension combines the transactions within household across category and hence removes the assumption of independence between categories. The models were built using both a pooled structure and a fixed effect structure, each within a non-hierarchical and hierarchical structure based on the panel id of the household as the hierarchical term.

8.1.2.1 RQ7: Pooled versus Fixed Effect Model structures

The diagnostics for the models discussed in the combined category model section suggest all four are delivering a statistically relevant representation of the underlying data. Table 41 below is a duplicate of the table discussed in the combined category model section.

	Pooled Non Hierarchical	Pooled Hierarchical	Fixed Effect Non Hierarchical	Fixed Effect Hierarchical
R-Squared (adj)	69.447%	72.202%	70.435%	72.858%
Mean Deviance	179,127	154,772	176,177	152,502
Penalty	42	1,564	63	1,584
DIC	179,169	156,336	176,240	154,087
MAPE	6.636%	6.207%	6.534%	6.126%
Variance (between purchases)	0.240	0.198	0.235	0.195
Variance (between housholds)		0.047		0.045
Variance (between household) t-stat (sig)		58.358(0)		57.881(0)
Variance Partition Coeficient		19.150%		18.766%
Rank of prefered model	4	2	3	1

Table 41: Model diagnostics comparison across all combined models

From Table 41, it can be seen the pooled models, where each parameter yielded one estimate across all four categories performed diagnostically similar to the fixed effect models. This implies that one coefficient estimate across all categories is almost as good a representation of the underlying data as the category specific estimates. The price elasticity of the four categories are similar (in fact statistically the same) and hence the same coefficient is a good means of estimating the four categories. Therefore, this may be a reason why the model is performing as well (almost) as the fixed effect model since the four categories in question have similar coefficients. This presumably would not generalise to all categories and if a category has a much different price elasticity for example, then the "average" estimated by the model would be of no use to any category and would lead to management making the wrong decision in all cases, underestimating the effect for some categories, whilst overestimating for others.

The fixed effect model provides the flexibility of accounting for any such significant differences between the coefficients of the parameters, the expense being the degrees of freedom demanded by the additional coefficients. In this case, where *n* is large, this is of little concern, however this is a factor which would be of consideration if the data had fewer degrees of freedom at its disposal. As mentioned briefly above, but worth pointing out again, the fixed effects model becomes complicated in terms of the number of offsets and how they all interact. For example, there is at some level an offset for a category effect crossed with the utilitarian reinforcement group, crossed with a supermarket own brand indicator and a level of informational reinforcement. This can be confusing to recreate and interpret which is a trade off from a statistically superior model versus a practical model to use and interpret. Also, the building of such a model with numerous offsets within a Bayesian model with informative priors is a further complication, since the nature of numerous levels of offsets has implications for a prior distribution, since the value of an offset parameter relies on what the base level may be. For relatively simple offset models this is much less of an issue, however models with three levels of offset has many moving parts and controlling for each of these offsets is a very difficult task. A conditional offset would help if such a thing existed, whereby the offset could be conditional on the posterior base value, though this is not something which has been seen in the literature.

In summary, answering RQ7 is complicated. From this specific study, it would be correct to state the pooled and fixed models are performing relatively the same, though logic does tend to warn this may not be a result to generalise to other categories. However, the fixed effects model does have its own challenges in terms of complexity both structurally and also interpretively. Taking these points into consideration, the study opts to prefer the fixed effect models since they offer more flexibility for category generalization and also the diagnostics are marginally better than those of the pooled models in both the non-hierarchical and hierarchical model cases.

8.1.2.2 RQ8: Differences in coefficients between combined and separate models

In considering RQ8, Table 48 below shows the diagnostic statistics for the non-hierarchical and hierarchical vague models across all four separate category model and for the combined category model.

	Biscuits		Fruit Juice		Yellow Fats		Beans		Combined	
	Non Hierarchical	Hierarchical	Non Hierarchical	Hierarchical	Non Hierarchical	Hierarchical	Non Hierarchical	Hierarchical	Fixed Effect Non Hierarchical	Fixed Effect Hierarchical
R-Squared (adj)	45.291%	55.863%	20.764%	55.185%	30.967%	58.119%	58.038%	76.697%	70.435%	72.858%
Mean Deviance	81,152	69,379	36,118	24,820	37,902	23,828	19,464	12,217	176,177	152,502
Penalty	18.2	1323.0	21.0	845.7	13.1	1244.0	14.1	763.6	63.1	1584.0
DIC	81170	70702	36139	25666	37915	25072	19478	12981	176240	154087
MAPE	5.93%	6.55%	6.24%	4.42%	6.05%	4.25%	5.84%	4.27%	6.53%	6.13%
Variance (between purchases)	0.221	0.182	0.318	0.187	0.201	0.127	0.243	0.143	0.131	0.125
Variance (between housholds)		0.039		0.144		0.076		0.105		3.601
between household t-stat (sig)		23.135(0)		19.118(0)		23.714(0)		18.67(0)		185.562(0)
Variance Partition Coeficient		17.582%		43.409%		37.428%		42.374%		96.632%

Table 42: Model diagnostics of combined and separate categories model diagnostics

The R-squared (adjusted) for the combined model shows the model is a better representation of the data for all instances of the non-hierarchical model and for all but the beans category for the hierarchical model. Also, philosophically, it would seem more logical that the assumption of independence between categories and within household be removed as household arguably have similar patterns of purchase, both from a price elasticity perspective and also from a psychological perspective with regards to informational and utilitarian reinforcement perspective. Also, attitudes to supermarket own label products may be similar across categories within a household. Hence a combined category model seems a favourable choice to recommend to management. However, it has been shown how the complexity of the model increases and this has implications both on the interpretation of parameters and on the computing power required to run the models (discussed in more detail below). However, for this study utilising four categories, this combined approach would be the recommendation.

Table 43 below shows the parameter estimates for both the separate and combined hierarchical vague models. The coloured boxes indicate the parameters which are statistically different from zero, blue being negatively so and red, positively so. The directional results within category are similar for the separate and combined models as discussed previously, hence the implication of removing the assumption of independence between categories has improved the fit of the model rather than changing the direction of the results.

	Biscuits				Fruit Juice			
	Sep	Sep CI	Comb	Comb CI	Sep	Sep CI	Comb	Comb CI
Log Price	-0.702	-0.71, -0.695	-0.696	-0.703, -0.688	-0.531	-0.549, -0.513	-0.524	-0.53, -0.519
Informational x Ut1	0.033	0.026, 0.039	0.032	0.025, 0.038	0.128	0.111, 0.145	0.161	0.156, 0.166
Informational x Ut2	0.054	0.046, 0.062	0.058	0.05, 0.066	0.023	-0.006, 0.052	0.025	0.017, 0.032
Supermarket Inf Ut1	-0.001	-0.008, 0.005	-0.001	-0.008, 0.006	-0.017	-0.03, -0.003	-0.034	-0.038, -0.029
Supermarket Inf Ut2	-0.081	-0.09, -0.071	-0.080	-0.089, -0.07	-0.037	-0.07, -0.003	-0.043	-0.052, -0.033
Christmas	0.043	-0.009, 0.094	0.059	0.007, 0.113	0.014	-0.045, 0.074	0.014	-0.039, 0.067
Christmas Ut1	-0.015	-0.094, 0.067	-0.018	-0.097, 0.06	0.080	-0.09, 0.258	0.079	-0.097, 0.06

	Yellow Fats				Baked Beans			
	Sep	Sep CI	Comb	Comb CI	Sep	Sep CI	Comb	Comb CI
Log Price	-0.448	-0.462, -0.435	-0.458	-0.463, -0.452	-0.476	-0.499, -0.453	-0.570	-0.578, -0.563
Informational x Ut1	0.157	0.146, 0.168	0.148	0.144, 0.153	0.029	0.014, 0.044	0.008	0.003, 0.013
Informational x Ut2	-0.001	-0.009, 0.007	-0.002	-0.007, 0.003	0.005	-0.008, 0.017	0.049	0.044, 0.054
Supermarket Inf Ut1	-0.033	-0.045, -0.021	-0.044	-0.048, -0.039	-0.088	-0.107, -0.07	-0.123	-0.128, -0.117
Supermarket Inf Ut2	0.004	-0.017, 0.025	0.022	0.014, 0.03	0.031	0.009, 0.054	-0.019	-0.026, -0.012
Christmas	-0.044	-0.085, -0.002	-0.068	-0.112, -0.023	0.010	-0.063, 0.08	-0.008	-0.062, 0.045
Christmas Ut1	0.030	-0.069, 0.13	0.033	-0.097, 0.06	0.133	-0.011, 0.277	0.158	-0.097, 0.06

Table 43: Summary of combined and separate parameter value

8.2 The BPM as a Measure of Brand Equity

Many authors have defined the brand to be of a physical representation e.g. the American Marketing Association's definition, cited in Kotler *et al.*, (1999, p. 442). "the name, term, sign, symbol or design or a combination of them intended to identify the goods and services of one seller or group of sellers and to differentiate them from those of competition". Murphy (2001) says it is the presentation which differentiates from competition.

However other authors have suggested there are other less tangible benefits, e.g. Dibb *et al.* (1997) say the brand may be differentiated through its physical characteristics or any other feature. Ambler *et al.* (1992) and Webster (1994) claim a brand is a bundle of consumer benefits which may be tangible or intangible, which hints at a psychological benefit. Style and Webster (1992) claim the brand claims the brand is responsible for elevating the consumer value in such a way that its sum of attributes have a higher value than the sum of the parts. Farquhar (1989) agrees stating the value is responsible for enhancing the value above the product's functional purpose.

These benefits are claimed to benefit both the firm (e.g. Aaker (1991), Bennett (1988), Dibb *et al.*, (1997), Kotler *et al.*, (1996), Watkins (1986)) and the customer (e.g. Aaker (1996), Ambler (1992), de Chernatony and McDonald (1992), Goodyear (1993), Keller (1993), Levitt (1962), Murphy (1990), Sheth *et al.*, (1991)).

Aaker (1996), Style and Ambler (1992) state the value to the organisation comes in the form of a brand equity which can be above and beyond that of the value of the product itself. The understanding of this equity can be valuable to the organisation to appealing to consumers' tangible and intangible desires of the purchase. The authors do not specifically state this as a specific psychological function.

From the behavioural based brand equity literature, the BPM helps to realise how the equity of brands may be quantified. Biel (1992) claims the equity lies beyond the physical products themselves. This is shown through the Behavioural perspective framework where the Informational reinforcement is a significant mediator of sales volume. There is no physical benefit associated with the Informational reinforcement and can only be a psychological benefit above and beyond what can be explained through price changes or through a Utilitarian reinforcement. This effect is not always a positive one and it has been demonstrated that higher Informational reinforcement brands have a negative impact on volume when associated with lower Utilitarian informational products and with supermarket own brand products. This is an interesting implication as it demonstrates the psychological association with higher equity is not always a positive one. This also suggests an interesting field of research for the subject of equity generally if in fact there are circumstances where lower equity products are more desirable to consumers psychologically than higher equity brands, even when the effect of price is accounted for. This effect is not consistent across all categories which may also require further research and more categories of research in order to determine any systematic insights which may contribute to knowledge.

Cobb-Walgren *et al.* (1995) explore equity with price premium products. The elasticity of demand coefficients across all four categories are inelastic (<1) suggesting the categories themselves may be stronger in equity. However, the significance of the Utilitarian and Informational reinforcement above and beyond the price elasticity demonstrates there is a psychological effect which can be theorised through a behavioural perspective.

Ailawadi et al, (2003) associate the effects associated with two similar products, though some of which are branded and some are not. However, Barwise (1993) suggest it is very difficult to estimate what each individual brand name may bring and exemplifies the difficulty of evaluating Coke if the name does not exist. The author agrees that attempting to allocate an

equity value to each separate brand is challenging and a quantitative model of this nature which is generalising across category is not relevant as a model of this nature.
However, when modelling the category in its entirety, statistically significant differences are seen for the supermarket own brand versus the branded products in terms of the Informational and Utilitarian reinforcement variables. These can prove insightful at the category level in terms of how products are managed either on-shelf or even within a manufacturer's portfolio strategy.

8.3 Incorporation of Bayesian practice within management

Often managers are faced with risk determination and require the ability to know to what degree a certain hypothesis may be true or not. Despite frequentist methods not being set up to directly answer the question, the values often are interpreted in this manner, usually due to ignorance of other alternatives (Bergerud and Reed 1998). Bayesian statistics is modelled through embracing uncertainty around events (O'Hagan, 1984). Probabilities are assigned to various possible outcomes, initially based on previous results or what a subject expert may expect to see. These probabilities are then updated in light of new data and hence truths are confirmed or myths exposed (Bergerud and Reed, 1998). With my personal experience in large organisations, I argue this form of risk management seems an almost idyllic way for managers/behavioural researchers or indeed organisations to learn.

For example, under a Bayesian paradigm managers may have answers to very practical questions such as "With posterior probability [x] the total volume of wood in the stand lies between 46 000 and 52 000 m3" or "the posterior probability of [the informational reinforcement parameter of the model being positive is y]" (Bergerud and Reed, 1998, p. 90, *[added by the author]*).

However, most organisations do not embrace the Bayesian culture and frequentist methods dominate management (Bergerud and Reed, 1998).

Several barriers seem to exist when it comes to Bayesian methods in management. Within academic institutions, statistics teaching is broadly around the frequentist paradigm and hence as these students mature into management roles they adopt the same frequentist methods

(Bergerud and Reed, 1998). Earlier in the 20th century, the nature of *subjectivity* within the Bayesian framework was regarded through the lens of Comte as “a capricious, arbitrary quality of the mind, responsible for not only inter- but also intra-individual differences” (Daston, 1994, p. 342). Bayesian techniques which are inverse probability models, were criticised in Fisher’s 1925 popular work *Statistical Methods for Research Workers which stated* “… the theory of inverse probability is founded upon an error, and must be wholly rejected.” (Fisher, 1925, p. 10). The inclusion of an equal prior also met with resistance (Edwards, 2004; Fienberg, 2006) with Newman stating the equal prior was “illegitimate” (Perks, 1947, p. 286).

Post WWII Bayesian ideas were, arguably paradoxically, severely restricted within statistical teaching and dominated by frequentist statisticians; paradoxically since Bayesian analysis had helped US forces capture the U-boat leading to the cracking of the Enigma code (McGrayne, 2012). However, so confidential was the nature of the work undertaken within the Bletchley Park, this work was never known to the public (Cabantous and Gond, 2015). As industry was rebuilding post WWII, next generation management which would lay the blocks of economic recovery were never exposed to the Bayesian way of thinking.

8.3.1 Engaging Management

Putman (2002) suggests the distinction between facts and judgments of stakeholders is a useful tool and distinctions can be made between them. Stakeholder involvement can also lead to benefits throughout any project from securing funding, identifying research questions and ensuring the research findings are better embraced. Models need to be seen through a worldview lens which can improve the end model and hence link the researchers with stakeholders who have specific insights or specialised knowledge about a project is beneficial (Welp *et al.,* 2006). Usually performed in smaller groups, these stakeholder discussions play a fundamental role in shaping organisational knowledge (Senge, 1990). The engagement of such stakeholders into the Bayesian process is essential if the paradigm is to be adopted within management, more so than frequentist methods given the input required for the prior distribution which, from the literature is the main difference associated with the Bayesian process.

Welp *et al.*, (2006) suggest the elicitation of quantitative data, facts and expert judgements is a means of capturing views of stakeholders at the start of a project, achieved through data mining or a formal elicitation process. This current study exemplifies this by using data analysis as a means of informing the prior distributions at the start of the study. However, this process is reliant purely on the calculation from the data without the contextual view of the stakeholder. The elicitation process allows the stakeholder to contribute to the calculation of any prior informative parameter to help gain acceptance of the process and project. Within Bayesian based studies, the vague prior dominates over informed priors, usually due to the difficulty in calculating an informed prior (Moala and O'Hagan, 2010). Indeed, this was seen to be the case with the current study especially when considering a combined four category model.

Within the Bayesian context, "*elicitation is the process of formulating a prior density about one or more uncertain quantities to represent a person's knowledge and beliefs*" (Moala and O'Hagan, 2010 p. 1635). In practice, there will always be some information available about the parameter, besides the data itself and the role of Bayesian inference is to help gain information around this prior distribution to help better understand the posterior distribution (O'Hagan, 1994).

Elicitation differs from other similar techniques. Discussion methods such as Delphi, for example, whereby views are discussed and iterations result in a consensus may suffer, as often the consensus is the view of the perceived expert rather than the view of the group (Aspinall, 2010). Within elicitation, experts are questioned to gain the information required. An "expert" in this case is defined as "a person who has background in the study area and enough knowledge to answer questions related to [the parameters in question]" (Moala and O'Hagan, 2010 p. 1636 *[added by the author]*).

Advantages of elicitation is the ability to gather a range of values about a parameter without the contributors having to understand the technical statistical theory as "experience in a subject matter is not the same as experience in statistics and probability" (Kadane and Wolfson, 1993, p. 3). It is assumed an expert is able to deduce some statistical summaries of the believed distribution, e.g. its mean, mode, moments, spread, limits etc., however in practice some training is usually required to achieve this (Alpert and Raiffa, 1982; Chaloner *et al.*, 1993).

The researcher then uses the same Bayesian process to obtain a posterior distribution for the parameter elicitation, i.e. the researcher can take the information given by the expert and infer a distribution of the expert's belief about the parameter. This becomes the likelihood and updates the researcher's initial belief about the parameter. The result becomes the posterior estimate of the elicitation parameter which can then be used as a prior to the model itself (Moala and O'Hagan, 2010).

There are limitations to the elicitation process. In the same way as temporal settings and historical accounts impact consumer behaviour (e.g. Foxall, 2013), the same is the case for the process of elicitation, in particular, the setting and context of the question (Oakley and O'Hagan, 2015). Psychologists' state heuristics may lead to biases in the expert's elicitated judgments, in particular an *availability bias* which leads an expert to bias their elicitation from their strongest associations or most recent experiences (especially bad experiences which tend to be more predominant in the mind) (Oakley and O'Hagan, 2015). Another common bias is *anchoring bias.* This is where an expert will condition their answers of subsequent related questions based on their response to an initial question, even if that question was not answered entirely correctly, in order to maintain the consistency of their answer thread (Oakley and O'Hagan, 2015).

Ways of dealing with this is the elicitation of a number of experts, independently to avoid consensus of the group (Aspinall, 2010; Cooke, 1991). When all views are gathered, this reduces (or highlights) any bias within interviews. Cooke (1991) also suggests asking questions of the expert where the truth is known to the researcher (though the expert is not aware of this). The expert opinions should then be weighted by their "performances" based on these known-truth questions.

There is a degree of subjectivity with the calculation of the prior as the researcher infers the posterior density of the elicitated parameter and (Moala and O'Hagan, 2010, p. 1636) point out the majority of the elicitation literature seem to result in a "convenient parametric" distribution.
Also, the process is complicated when multiple parameters is required to be elicitated within a model especially if the parameters are not independent and hence elicitation around covariance parameters is required (Moala and O'Hagan, 2010). Experts are seldom capable of

elicitating second order moments (Kadane and Wolfson, 1998). This impacts the combined category model where offsets complicate the model.

Andrews and Baguley (2013) argue the choice of prior distribution is a part of the modelling assumptions, and like other (more general) model assumptions, may be a good or bad choice and may need to be revisited or changed (Andrews and Baguley, 2013). Andrews and Baguley (2013) claim the field of psychology needs to use the range of tools provided by both the frequentist and Bayesian methods to help solve the complex problems faced in psychology. This echos the views of Efron (2005). The choice of model, functional form and assumptions around a statistical model will always be incomplete and always contain a degree of uncertainty. As Macdonald (2002, p. 187) wrote: "if the incompleteness of probability models … were more widely appreciated psychologists and others might adopt a more reasonable attitude to statistical tests, the debate about statistical inference might die down, and the emphasis could shift toward better understanding and presenting data".

8.3.2 Resources required to run of Bayesian analysis

Further considerations must be taken into account when deciding to run Bayesian analysis. These lie outside of the paradigm argument and are more functional issues. Nonetheless, they are important to take into account and can be seen as further barriers to the adoption of Bayesian analysis.

8.3.2.1 Run time

A consideration of the decision to run a Bayesian resides in the time it takes to estimate the parameters. MCMC chains can take time to converge and then further iterations are required to estimate the parameters. As an example the estimation of the fixed effects hierarchical models in this study required almost 23 hours of estimation time running through the University's desktop PC (which one would consider to be a of a good computational power). One model crashed the system twice and hence required re-running. This information is not intended as a basis of aggravation, merely the frustration an analyst may experience if (s)he was under pressure to produce an analysis for management for a given deadline. This does have a commercial implication to the Bayesian philosophy and it may require a step change in the speed of complex models estimation before this means of analysis of complex models truly becomes mainstream and alternative to frequentist methods.

8.3.2.2 Software of Bayesian analysis

This study has uncovered different means of software to run the Bayesian estimations. Initially, WinBUGS software was employed which is prevalent in the Bayesian literature. WinBUGS is a good platform to run the analysis with a very logical approach whereby the model is defined, the data is defined, initial values can be loaded, the number of iterations is defined, the burn in is defined the model is compiled and finally the model can be run. At each stage the user can monitor how each stage is run and is transparent in the software. The parameters can be monitored and exported from the software, as can the parameters and predicted values. Issues were, however experienced when exporting the estimates of the model given the complexity and also the large number of estimates to be exported.
This led to the exploration of the Rjags software. The experience here was a little different. The process was more complicated to author and less intuitive to create. However, the benefit is the model then resides within the R environment. This means the outputs of the model can be saved within the R environment which makes the manipulation of the data more stable and also other analytical and graphical tools are available to utilise on the data through the R environment. Also, exporting the data to other software is more stable.

The issue being raised is the lack of commercial “off the shelf” or “point and click” packages which exist for frequentist methods. This does have a non-trivial bearing on the pool of recruitment talent which exists and also on to training budgets of organisations if analysts require increased training on the more complex and less user friendly software options.

8.4 Future Considerations and limitations

8.4.1 Functional Form

The hierarchical models were run with random effects intercepts based on the household panel id. This is the simplest form of the random effects hierarchical model. It was also chosen as the household id was to be a representation of the wider population of British households (Field *et al.*, 2012). An extension of this model would be the introduction of hierarchical slopes to the model. These would be hierarchical terms around the focal parameters and in effect would allow households to have varying slopes for each parameter. This would generally give a better predictive model as any delta changes in the parameter estimates for a household would result in an improved model, in the same way as the inclusion of the hierarchical intercept improved the models in this study. There are issues, however, when both the number of households is many as in this case. The hierarchical term could be that of a fixed effect, whereby each household *h*-1 (hence 1689-1=1688) would have its own fixed effect delta from a base household. Given the large number of households, this does drastically increase the number of degrees of freedom required, especially if the hierarchical framework was given to all seven focal parameters. Another form of hierarchical measure would be a random slope, whereby the variance of the parameter is represented by one variable. This reduces the number of degrees of freedom to one per parameter (as seen for the random intercept in the current model). However, the way in which this is coded within the Bayesian model requires an underlying sub-loop of the 1688 households (*h*-1). Hence, despite the degrees of freedom being unaffected, the model would need to monitor and develop MCMC chains for the 1688 nodes per parameter. Given the current complexity (circa 23hrs to run the combined hierarchical model), introducing a further number of these random slopes may be a barrier to being able to run the model at all, or at least in a feasible time frame. Hence this is a limitation of the Bayesian estimation rather than a limitation of the hierarchical structure.

A limitation to hierarchical models would be the ability to predict an additional household. Within a non-hierarchical structure, a prediction could be made from the model if the independent variable is known and this is a reasonable requirement to gaining a model predictive score. Presumably if a product resided in one of the four categories, with a price, an informational and utilitarian reinforcement value, an indication of whether or not it is a supermarket own brand and also whether it is the Christmas week or not, a simple calculation can predict purchase level. However, within a heretical structure where household id represents the hierarchy, the household would also be required in order to include this within the prediction. This may or may not be possible or even relevant to what the manager is requiring, however this does need to be fed into the model in order to adjust the hierarchical intercept of the model.

These issues need to be addressed when considering model functional form and the trade-off of what can be achieved.

8.4.2 Informative Priors

The informative nature of the priors for the separate models was calculated using a calibrated means. It would be useful to also move to an elicitation model whereby stakeholders and experts can unite to build the informative parameter. I would argue at the very least this demonstrates the diversity/unity of views within the organisation. It also brings stakeholders into the modelling process, gaining buy-in to the project. Stakeholder meetings are common in my professional experience and are a great means of engaging stakeholders as Bergerud and Reed (1998) claim, and the extension of involving them at a model build phase can only build on this. The current study had many data points and the categories were developed by a world FMCG panel data, hence complete and reliant data. For other sorts of data this may not be the case and the importance of elicitation becomes much more important, if only to control for absurd situations or benefit from wider more robust studies. Being able to do this at the start of the process and mathematically incorporate the results into the model is a true benefit of the Bayesian process I would argue.

8.4.3 Consumer setting

The Skinner-based research undertaken initially focussed on animal behaviour within a laboratory style setting where specific behaviour can be monitored by controlling for any other impact due to the environment or situational setting. Hence the behaviour being observed was as independent of other (non-cognitive) factors as can be realistically hoped for. When considering consumer behaviour, situational context needs to be taken into account (e.g. Foxall, 2010) since it is not realistic to assume the environment or setting of the behaviour has no impact on consumer choice. The BPM accommodates this through its open to closed continuum consumer setting (as discussed in the literature review). Within this study, the consumer setting is relatively open. Yan (2012) opted to decompose the nature of the UK supermarket into ones which were more open or closed depending on the size of the supermarket (larger supermarkets were assumed to be more open due to more browsing time and more items to browse). This may be an area to investigate in further research, especially when considering the supermarket own brands. The double jeopardy effect may be more relevant to larger supermarkets' brands than to smaller supermarkets which may influence Utilitarian and/or Informational reinforcement of products.

8.4.4 Marketing Mix Variables

It is important to note this study is being carried out at a category level, whereby evidence of economic, psychological and marketing effects within the BPM are being observed at a category level. This is useful in considering the behaviour of the consumer towards categories (what this study aims to do). It is not meant to be an organisation based tactical study of marketing mix implementation. However, this could be a further area of research whereby the incorporation of marketing mix variables is considered. These could include (but not limited to) off shelf and on-shelf displays within supermarkets, advertising channels, in store sampling or tasting events or sponsorship events.

The construction of the model may therefore be the consideration of items at the brand (or even SKU) level. At an SKU level, changes in the marketing mix of a SKU can be accurately measured in terms of its impact on that specific SKU when SKU specific promotions or price changes are implemented. The resulting uplift multipliers can be observed. Often, other SKUs within the same brand are impacted of course (e.g. a price promotion on a two pack may have negative consequences for the one pack within the same SKU). Also, the promotion may cannibalise other brands and these brands may or may not belong to the same overall brand owners, hence the promotions may have consequences for the category management of brand ranges.

Advertising and sponsorship may not have SKU specific effects though could affect brand performance and hence cannibalise other brands, again within (or not) the same brand owners' portfolio. However, this may result in a static category whereby volume share mix changes within brand though the category remains flat.

This much increases the complexity of the model and a compromise would probably be required in terms of the number of brands (or SKUs) included in such a study. This is a wider issue with model based studies of this nature as to the knowledge being sought versus the complexity which can be realistically built. It has already been seen the increase in

computational power required for the Bayesian approach to work and this added complexity will fuel this further. Also, if informative priors based models are required then more is required to determine these ongoing.
However, the BPM structure offers this form of model build and, as most studies are, the limitation is usually within the design, computational restrictions and complexity of interpretation, rather than the theoretical framework.

8.4.5 Geographical and Category Limitation

The study is based on four UK FMCG categories. The positivism epistemology underlying this study suggests a set of rules may be established which can be generalised. The author acknowledges the rules are intended for generalisation to each category though limitations exists in terms of whether other categories would perform in a similar fashion or whether the same categories within different geographical areas would behaviour in the same way. This may be an aspect of further research.

8.4.6 Timeframe of the Data

The data used in this study are sourced from a household scanner data, relating to four FMCG categories within Great Britain. The categories within the studies offer the characteristics of being sufficiently different in terms of product type; though sufficiently similar to be purchased within a typical shopping basket. This becomes useful in the study when assessing the relevance of combining the categories into one behavioural model across all four categories.

The data relates to the time period of week ending 17 July 2004 to 9 July 2005. These data are therefore eleven years old meaning there will be a significant limit to the insight that can be drawn from the data directly relating to the brands and products within this study. Some of these brands may have evolved over the eleven years; through brand extensions, differing levels of investment, brand positioning and discontinuation. In this light, the aim of this study is not to offer insights based on these products which would be relevant to a manufacture today. The aim of the study, is, however, to develop and offer methodology which may be used to explore similar data types from any year of study. The Behavioural Perspective

Model's theoretical framework has been relevant to describing the behaviour of a consumer over many years (Foxall and James, 2003; Foxall and Schrezenmaier, 2003; Foxall *et al.,* 2004; Foxall *et al.,* 2006; Oliveira-Castro *et al.,* 2005; Romero *et al.,* 2006) and this study builds on this through the inclusion of supermarket own brand and seasonality variables situated within the BPM Informational and Utilitarian framework. Furthermore, the demonstration of how the Bayesian hierarchical framework can be applied through the BPM demonstrates how this methodology provides more tools to better understand consumer behaviour.

The thesis, therefore, demonstrates contribution to the field through the inclusion of supermarket own brand indicators and seasonal variables, crossed with the BPM Informational and Utilitarian variables; structure of the models (hierarchical vs non-hierarchical) and the estimation through Bayesian inference (using both vague and informed priors). This is all constructed through the BPM framework, demonstrating the flexibility of this tried and tested theoretical framework.
As with much research, the timely conclusions specific to the brands and products within the research nay become dated, however the methodology and theoretical approach contributes to knowledge advancement, especially within the BPM framework.

8.4.7 Portfolio and Segmentation Limitations

It is also worth discussing the portfolio limitations of the study. The data has not assessed the distribution limitations of the categories and not taken into account that some organisations may segment their portfolio of products in such a way whereby some brands are more widely distributed than others.
Also, it is worth noting the extent to which stores try and influence consumer behaviour in the form of atmospheric conditions (Turley and Milliman, 2000), background music (Areni and Kim, 1993), shelf space allocation (Reyes and Frazier, 2007) and supermarket own brand shelf space management (Nogales and Suarez, 2005). These aspects were not considered during this study.

References

Abelard. *Cause, chance and Bayesian statistics.* Available at http://www.abelard.org/briefings/bayes.htm [Last accessed 03/01/2016].

Aaker, D. A. (1991). *Managing brand equity*. New York: The Free Press.

Aaker, D. A. (1992). The value of brand equity. *Journal of business strategy, 13(4)*, pp. 27-32.

Aaker, D. A. (1996), Building Strong Brands, New York, The Free Press.

Aaker, D. A. and Jocobson, R. (1994). The financial information content of perceived quality. *Journal of Marketing Research, 31(2)*, pp. 191–201.

Aaker, D. A., and Jacobson, R. (2001). The value relevance of brand attitude in high-technology markets. *Journal of marketing research, 38(4)*, pp. 485-493.

Ailawadi, K. L., Lehmann, D. R., and Neslin, S. A. (2003). Revenue premium as an outcome measure of brand equity. *Journal of Marketing, 67(4)*, pp. 1-17.

Ajzen, I. and Fishbein, M. (1980). *Understanding attitudes and predicting social behaviour*, Englewood Cliffs, NJ: Prentice Hall.

Ajzen, I. (1985). "From intentions to actions: a theory of planned behaviour", In: Kuhl, J. & Beckman, J. (Eds.), *Action Control: From cognition to behaviour*, Berlin: Springer Verlag, pp. 11-39.

Allen, D. 1989. The Financial Management of Brands. *Management Accounting, March*. Pp. 20–21.

Allen, C. T. and Madden, T. J. (1985). A closer look at classical conditioning. *Journal of Consumer Research* 12, pp. 301–315.

Allen, C. T., and Janiszewski, C. A. (1989). Assessing the role of contingency awareness in attitudinal conditioning with implications for advertising research. *Journal of Marketing Research*, 26(1), pp. 30–43.

Allen, D. (1990). Creating value - The financial management of brands. *CIMA.* Available at:http://hdl.handle.net/10068/507293

Alpert, M., Raiffa, H. (1982). A progress report on the training of probability assessors. In: Kahneman, D., Slovic, P., Tversky, A. (Eds.), Judgement Under Uncertainty: Heuristics and Biases. Cambridge University Press, Cambridge, pp. 294–305.

Ambler T (1992), "Need-to-Know-Marketing", Century Business, London

Andrews, M., and Baguley, T. (2013). Prior approval: The growth of Bayesian methods in psychology. *British Journal of Mathematical and Statistical Psychology, 66(1)*, pp. 1-7.

Atkin, C.K. (1984). 'Consumer and social effects of advertising'. In: B. Dervin and M.J. Voigt (eds.), Progress in communication sciences. Norwood, NJ: Ablex.

Areni, C. S., and Kim, D. (1993). The influence of background music on shopping behavior: classical versus top-forty music in a wine store. *ACR North American Advances.*

Aspinall, W. (2010). A route to more tractable expert advice. *Nature* 463(21), pp. 294-295.

Badot, O., Carrier, C., Cova, B., Desjeux, D. and Filser, M. 2009. The contribution of ethnology to research in consumer and shopper behaviour: toward ethnomarketing. *Research et applications en marketing 24(1)*, pp.93-111.

Bagozzi, R. P. (2000). The poverty of economic explanations of consumption and an action theory alternative. *Managerial and Decision Economics* 21, pp. 95–109.

Baldinger, A. L. (1990), Defining and Applying the Brand Equity Concept: Why the Researcher Should Care. *Journal of Advertising Research, 30* (June–July), RC2–RC5.

Baldinger, A. L. (1992). What CEOs Are Saying About Brand Equity: A Call to Action for Researchers. *Journal of Advertising Research, 32* (July–August), RC6–RC12.

Baldinger, A. L. and Rubinson, J. (1997). The jeopardy in Double Jeopardy. *Journal of Advertising Research* November, pp. 37-49.

Baldinger, A. l., Blair, E. and Echambadi, R. (2002). Why brands grow. *Journal of Advertising Research*, pp. 7-14.

Bales, M. (2009). Marketing and minds in the early 20th century: How psychology enhanced advertising, *The Harvard Brain*. [http://www.musingsonmind.org/featured/Bales_Marketing]

Bandyopadhyay, S., Gupta, K. and Dube, L. (2005). Does brand loyalty influence double jeopardy? A theoretical and empirical study. *Journal of Product and Brand Management* 14(7), pp. 414–423.

Barker, R.G. (1987). `Prospecting in environmental psychology: Oskaloosa revisited", in Stokols, D. and Altman, I. (Eds), Handbook of Environmental Psychology, John Wiley and Sons, New York, NY, pp. 1413-32.

Barwise, P. (1993). Brand equity: snark or boojum?. *International Journal of Research in Marketing*, *10(1)*, 93-104.

Bass, F. M., and Wittink, D. R. (1975). Pooling issues and methods in regression analysis with examples in marketing research. *Journal of Marketing Research*, pp. 414-425.

Bassi, F. (2011). The Dirichlet model: analysis of a market and comparison of estimation procedures. *Marketing Bulleting* 22, Technical Note 1.

BBC. (2013). *Yahoo to buy Tumblr for $1.1bn*. Available at: http://www.bbc.co.uk/news/business-22591026 [Last accessed 20/08/2013].

Bennett, P.D., ed. (1988). Dictionary of Marketing Terms, *Chicago: American Marketing Association.*

Bergerud, W. A. and Reed, W. J. Bayesian Statistical Methods (1998) pp. 89-104 in Sit, V., & Taylor, B. (1998). *Statistical methods for adaptive management studies*. British Columbia, Ministry of Forests Research Program.

Bernardo, J. M. (1999). Bayesian methods in the sciences. *Special Issue of Rev. Acad. Cien Madrid 93(3).*

Bernado, J. M. and Smith, A. F. M. (2000). *Bayesian Theory*. Chichester: Wiley.

Bettman, J.R. (1979). *An information processing theory of consumer choice*. Reading, MA: Addison-Wesley.

Biel, A. L. (1992), How Brand Image Drives Brand Equity. *Journal of Advertising Research, 32* (November), RC6-RC 12.

Bogart, L., and Lehman, C. (1973). What makes a brand name familiar?. *Journal of Marketing Research*, pp. 17-22.

Broadbent S. R. (1980). Price and advertising: Volume and profit. *Admap*, 532–40.

Brown, S. (1993), Postmodern marketing. *European Journal of Marketing, 27(4)*, pp. 19-34.

Browne, W. and Rasbash, J. (2004). *New developments in modeling, multilevel modelling* In: Hardy, M and Bryman, A. eds. *Handbook of Data Analysis*. London: Sage.

Bryman, A. (2008) Social Research Methods (Third Edition), New York: Oxford.

Buckley, K. W. (1982). The selling of a psychologist: John Broadus Watson and the application of behavioral techniques to advertising. *Journal of the History of the Behavioral Sciences, 18*, pp. 207–221.

Cabantous, L. and Gond, J-P. (2015). The Resistible Rise of Bayesian Thinking in Management: Historical Lessons From Decision Analysis. Journal of Management, 41(2), pp. 441- 470

Cameron, A. C., and Trivedi, P. K. (2010). *Microeconometrics using stata (Vol. 2)*. College Station, TX: Stata press.

Chaloner, K.M., Church, T., Louis, T.A., Matts, J.P. (1993). Graphical elicitation of a prior distribution for a clinical trial. Statistician 42, pp. 341–353.

Chang, S. W. (2007). *Behavioral economics: An empirical investigation of consumer buying behavior and price responsiveness in the British biscuit market* (Unpublished doctoral dissertation). Cardiff University.

Chaudhuri, A. (1995). Brand equity or double jeopardy? *Journal of Product and Brand Management*. 4(1) pp. 26 – 32.

Chauduri, A. (1999). Does brand loyalty mediate brand equity outcomes? *Journal of Marketing Theory and Practice, 2,* pp. 136-146.

de Chernatony, L. and Malcolm H.B. McDonald, M. H. B. (1993). *Creating Powerful brands*. Butterworth-Heinemann, Oxford.

Cobb-Walgren, C.J., Cynthin A.R. and Donthu, N. (1995) Brand Equity, Brand Preference, and Purchase Intent. *Journal of Advertising, 24*(3), pp. 25-40.

Colley, R.H. (1961). *Defining advertising goals for measured advertising results*. New York: Association of National Advertisers.

Colombo, R. A. and Morrison, D. G. (1989). A brand switching model with implications for marketing strategies. *Marketing Science 8(1)*, pp. 89-99.

Congdon, P. (2003). *Applied Bayesian modelling.* London: Wiley.

Cooke, R. M. (1991). *Experts in Uncertainty: Opinion and Subjective Probability in Science*. Oxford: Oxford University Press.

Daston L. J. (1994). How probabilities came to be objective and subjective. *Historia Mathematica, 21*, pp. 330-344.

Davies, J. D., Foxall, G. R. and Pallister, J. G. (2002). Beyond the intention-behaviour mythology: An integrated model of recycling. *Marketing Theory* 2, pp. 29–113.

Davis, R. A. (2001). A cognitive-behavioral model of pathological Internet use. *Computers in Human Behavior* 17(2), March, pp. 187–195.

Decaluwé, V. and Braet, C. (2005). The cognitive behavioural model for eating disorders: A direct evaluation in children and adolescents with obesity. Eating Behaviors 6(3), June, pp. 211–220.

Dibb, S., Simkin, L., Pride, W. M., Ferrell, O. C. (1997), Marketing: Concepts and Strategies, 3rd European ed., Houghton Mifflin, Boston, MA.

DiClemente, D. F. and Hantula, D. A. (2003). Applied behavioral economics and consumer choice. *Journal of Economic Psychology* 24, pp. 589–602.

Donovan, R. (2011). Theoretical models of behaviour change. In G. Hastings, K. Angus, & C. Bryant (Eds.), The SAGE handbook of social marketing (pp. 15–31). Thousand Oaks, CA: Sage.

Driel H.V., Nadall, V. and Zeelenberg, K. (1997). The demand for food in the United states and the Netherlands: a systems approach with the CBS model. *Journal of Applied Economics* 12, pp. 509-532.

Driver J.C. and G.R. Foxall, (1984). *Advertising policy and practice*. London: Holt, Rinehart and Winston; New York: St Martin's Press.

Duncan, C., Jones, K. and Moon, G. (1996). Health related behaviour in context: A multilevel approach. *Sot', Sci. Med.* 42(6), pp. 817-830.

Dunson, D. B. (2001). Commentary: Practical advantages of Bayesian analysis of epidemiologic data. *American Journal of Epidemiology, 153(12)*, 1222–1226.

Easterby-Smith, M., Thorpe, R. and Jackson, P. (2011) *Management research (Third Edition),* London: Sage.

Edwards W. (2004). Comment on Bellhouse, David R. The Reverend Thomas Bayes FRS: A biography to celebrate the tercentenary of his birth. *Statistical Science, 19*, pp. 34-37.

Efron, B. (2005). Bayesians, frequentisits and scientists. *Journal of the American Statistical Association* 100(469), pp. 1-5.

Ehrenberg, A.S.C. (1969). Towards an integrated theory of consumer behaviour. *Journal of the Market Research Society* 11, 305-337.

Ehrenberg, A. S. C. (1972). *Repeat-buying: theory and applications.* North-Holland

Ehrenberg, A.S.C. (1974). Repetitive advertising and the consumer. *Journal of Advertising Research* 14, 25-34

Ehrenberg, A.S.C. and Goodhardt, G.J. (1980). *How advertising works*. New York: J. Walter Thompson/MRCA.

Ehrenberg, A. S. C. and Goodhardt, G. (2002). Double Jeopardy revisited again, Marketing Learnings 7, *The Southbank University*, pp. 1-2.

Ehrenberg, A. S. C., Goodhardt, G and Barwise, P. T. (1990). Double Jeopardy revisited. *Journal of Marketing* 54, pp. 82-91.

Engel, J.F. and R.D. Blackwell, (1982). *Consumer behavior. (4th ed)*. Hindsdale, IL: Dryden.

Engel, J. F., Blackwell, R. D. and Miniard, P. W. (1995). *Consumer Behaviour, 8th Ed.* Chicago: Dryden Press.

Erdem, T., & Swait, J. (1998). Brand equity as a signaling phenomenon. *Journal of consumer Psychology, 7(2)*, 131-157.

Estes, W. K., Newell, A., Anderson, J. R., Brown, J. S., Feigenbaum, E. A., Greeno, J., Hayes, P. J., Hunt, E., *Society, 18*(60), pp. 1-15. Society, 33, pp. 237-240.

Fader, P. S. and Schmittle, D. C. (1993). Excess behavioral loyalty for high-share brands: Deviations from the Dirichlet model for repeat purchasing. *Journal of Marketing Research* 30(4), pp. 478-493.

Fairburn, C. G., Shafran, R. and Cooper, Z. (1999). A cognitive behavioural theory of anorexia nervosa. Behaviour Research and Therapy 37(1), January, pp. 1–13.

Farquhar, P. H. (1989). Managing brand equity. *Marketing research, 1(3).*

Fazio, R. H. (1986). How do attitudes guide behaviour. *Handbook of motivation and cognition: Foundations of social behavior, 1*, pp. 204-243.

Feyerabend, P. (1975). *Against Method*. London: NLB.

Field, A., Miles, J. and Field, Z. (2012). *Discovering statistics using R.* London: Sage.

Fienberg S., E. (2006). When did Bayesian inference become "Bayesian"? *Bayesian Analysis 1*, pp. 1-40.

Finley, A. O. (2013) Using JAGS in R with the rjags package

Fisher, R. A. *(*1925*). Statistical Methods For Research Workers*. Edinburgh, UK*: Oliver and Boyd.*

Foster, G. 1989. There's No Accounting for Brands. *Management Today, October,* pp. 90–96.

Foxall, G.R. (1983). Consumer choice. London: Macmillan; New York: St Martin's Press.

Foxall, G.R. (1984). Evidence for attitudinal-behavioural consistency: implications for consumer research paradigms. Journal of Economic Psychology 5, pp. 71-92.

Foxall, G. R. (1986a). Consumer theory: some contributions of a behavioral analysis of choice. *Management Bibliographic and Reviews* 12, pp. 27-51.

Foxall, G. R. (1986b). The role of radical behaviorism in the explanation of consumer choice. *Advances in Consumer Research* 13, pp. 187-191.

Foxall, G. R. (1986c). Theoretical progress in consumer psychology: The contribution of a behavioural analysis of choice. *Journal of Economic Psychology*, 7(4), pp. 393–414.

Foxall, G. R. (1987). Radical behaviourism and consumer research. *International Journal of Research in Marketing* 4(2), pp. 111-129.

Foxall, G. R. (1990/2004). Consumer Psychology in Behavioral Perspective. London and New York: Routledge. (Reprinted 2004 by Beard Books, Frederick, MD.)

Foxall, G. R. (1993). A behaviourist perspective on purchase and consumption. European Journal of Marketing 27(8), pp. 7–16.

Foxall, G. R. (1997). Marketing Psychology: The Paradigm in the Wings. London: Macmillan.

Foxall, G. R. (1999a). The behavioral perspective model consensibility and consensuality. *European Journal of Marketing 33(5/6)*, pp. 570–596.

Foxall, G. R. (1999b). The substitutability of brands. Managerial and Decision Economics, 20(5), 241-257

Foxall, G. R. (2001). Foundations of consumer behaviour analysis. *Marketing Theory* 1(2), pp. 165-199.

Foxall, G. R. (2002). Marketing's attitude problem – and how to solve it. *Journal of Customer Behaviour* 1(1), pp. 19–48.

Foxall, G. R. (2003). The behavior analysis of consumer choice: An introduction to the special issue, *Journal of Economic Psychology* 24, pp. 581–588.

Foxall, G. R. (2005). *Understanding Consumer Choice* London and New York: Palgrave Macmillan.

Foxall, G. R. (2010). *Interpreting Consumer Choice: The Behavioral Perspective Model.* New York: Routledge.

Foxall, G. R. (2013). Brand-related and situational influences on demand elasticity. *Journal of Business Research* 66, pp. 73-81.

Foxall, G. R. (2016a). *Addiction as Consumer Choice: Exploring the Cognitive Dimension.* London and New York: Routledge.

Foxall, G. R. (2016b). *Perspectives on Consumer Choice. From Behavior to Action, from Action to Agency.* London and New York: Palgrave Macmillan.

Foxall, G. R. and James, V. K. (2003). The behavioural ecology of brand choice: How and what do consumers maximize? Psychology and Marketing 20, pp. 811-836.

Foxall, G. R. and Schrezenmaier, T. C. (2003). The behavioural economics of brand choice: Establishing a methodology. *Journal of Economic Psychology* 25(5), pp. 675-695.

Foxall, G. R., Oliveira-Castro, J. M. and Schrezenmaier, T. C. (2004). The behavioural economics of consumer brand choice: patterns of reinforcement and utility maximization. *Behavioural Processes* 66(3), pp. 235-260.

Foxall, G. R., Oliveira-Castro, J. M., James, V. K., Yani-de-Soriano, M. and Sigurdsson, V. (2006). Consumer behaviour analysis and social marketing: the case of environmental conservation. *Behaviour and Social Issues* 15(1), pp. 101-124.

Foxall, G.R., Oliveira-Castro, J.M., James, J.K. and Schrezenmaier, T.C. (2011). Consumer behavioural Analysis I. The Behavioural Perspective Model. *Durham Research online*, pp. 1-9.

Foxall, G. R., Yan, J, Oliveira-Castro, J. M. and Wells, V. K. (2013). Brand-related and situational influences on demand elasticity. *Journal of Business Research* 66, pp. 73–81.

Frost, R. O. and Hart, T. L. (1996). A cognitive-behavioral model of compulsive hoarding. *Behaviour Research and Therapy* 34(4), April, pp. 341–350.

Gabor A. (1988). *Pricing*. London: Heinemann.

Geertz, C. (1973). *The interpretation of cultures.* New York: Basic books.

Gelman, A. (2007). Struggles with survey weighting and regression modeling. *Statistical Science 22(2)*, 153–164.

Gelman, A. (2010). Bayesian statistics then and now. *Statistical Science 25(2)*, pp. 162–165.

Gelman, A. and Rubin, D. B. (1992). A single series from the Gibbs sampler provides a false sense of security. *Bayesian Statistics 4*, pp. 625–631.

Goldsmith, R. E. (2004). Current and future trends in marketing and their implications for the discipline. *Journal of Marketing Theory and Practice* 12(4), pp. 10–17.

Goodhardt, G. J., Ehrenberg, A. S. C. and Chatfield, C. (1984). The Dirichlet: a comprehensive model of buying behaviour, *Journal of the Royal Statistical Society A* 147(5), pp. 621-655.

Goodyear, M. (1993). Reviewing the concept of brands and branding. *Marketing and Research Today, 21(2),* p. 75.

Gorn, G. J. (1982). The effects of music in advertising on choice behavior: A classical conditioning approach. *Journal of Marketing,* 46(1), pp. 94–101.

Hansen, F. (1976). Psychological theories of consumer choice. *Journal of Consumer Research 3*, pp. 117-142.

Hansen, M. H., Perry, L. T. and Reese, C. S. (2004). A Bayesian operationalization of the resource-based view. *Strategic Management Journal 26*, 1279–1295.

Hanson, D. and Grimmer, M. (2007) The Mix of Qualitative and Quantitative research in Major Marketing Journals, 1993-2002. *European Journal of Marketing, 41(1-2),* pp. 58-70.

Hantula, D. A., DeNicolis Bragger, J., & Rajala, A. (2001). Slips and falls in stores and malls: Implications for community-based prevention. *Journal of Prevention and Intervention in the Community, 22*, pp. 67–80.

Hart, S. and Murphy, J. (Eds) (1998), *Brands: The New Wealth Creators*, New York University Press, New York, NY

Hermstein, R. J, (1970). On the law of effect. *Journal of the Experimental Analysis of Behavior* 13, pp. 243-266.

Herr, P. M. and Russell H. F. (1988). The Attitude-to-Behavior Process: Implications for Consumer Behavior, *working paper, Indiana University*

Hirschman, E. C. (1986). Humanistic inquiry in marketing research: Philosophy, method and criteria. *Journal of Marketing Research* 23, pp. 237-249.

Hirschman, E. C. and Holbrook, M. B. (1992). *Postmodern consumer research: the study of consumption as text.* London: Sage publications.

Howard, J. A. (1983). Marketing theory of the firm. *Journal of Marketing* 47, pp. 90-100.

Howard, J. A. and Sheth, J. N. (1969). *The theory of buyer behaviour*. New York, NY: John Wiley & Sons.

Hunt, S. D. (1991). *Modern marketing theory. Critical issues in the philosophy of marketing science*. Ohio: South Western Publishing.

Jackson, T. (2004) cited in Bruce. M and Kratz, C. “Competitive Marketing Strategies of Luxury Fashion Companies” (2007) ” In: Hines, T and Bruce, M (eds) Fashion Marketing 2nd Edition. Elsevier/Butterworth-Heinemann. Oxford UK.

Jeffreys, H. (1931). *Scientific Inference*. Cambridge University Press.

Johnson, P. and Duberley, J. (2011). *Understanding management research*. London: Sage.

Jones, G. and Morgan, N. J. (1994). *Adding Value Brands and Marketing in food and Drink*. London: Routledge.

Kadane, J.B. and Wolfson, L.J., (1998). Experiences in elicitation. *Statistician 47*, pp. 3–19.

Kagel, J. H., Battalio, R. C. and Green, L. (1995). *Economic Choice Theory: An Experimental Analysis of Animal Behavior*. Cambridge: Cambridge University Press.

Kamakura, W. A. and Russell, G. J. (1993), Measuring Brand Value with Scanner Data. *International Journal of Research in Marketing,* I O(March), pp. 9-22.

Kassarjian (Eds.), Handbook of Consumer Behavior (pp. 124-1611. Englewood Cliffs, NJ: Prentice-Hall.

Kellaris, J. J. and Cox, A. D. (1989). The effects of background music in advertising: A reassessment. *Journal of Consumer Research, 16,* pp. 113–118.

Kellett, S. and Bolton, J. V. (2009). Compulsive Buying: A Cognitive–Behavioural Model Clin. Psychol. *Psychother* 16, pp. 83–99.

Keller, K. L. (2001). Building customer-based brand equity: A blueprint for creating strong brands.

Keller, K.L. (1993). Conceptualizing, measuring, and managing customer-based brand equity. *Journal of Marketing 57, pp. 1-22.*

Koop, G., Poirier, D. J., and Tobias, J. L. (2007). *Bayesian econometric methods*. Cambridge University Press.

Kotler, P., Armstrong, G., Saunders, J. and Wong, V. (1999). *Principles of Marketing 2nd European Edition*, Prentice Hall Inc. , New Jersey

Kuhn, T. S. (1970). *The structure of scientific revolutions (Second edition)*: University of Chicago Press.

Laing, R.D. (1967). *The politics of experience and the birds of paradise*. Harmondsworth: Penguin.

Lane, E. (2013) Horsemeat scandal. Can brands like Findus recover? http://www.bbc.co.uk/news/business-21413966

Lasser, W., Mittal, B. and Sharma, A. (1995). Measuring customer-based brand equity. *Journal of Consumer Marketing, 12(4):* pp. 11–19

Lavidge, R.J. and Steiner, G.A. (1961). A model for predictive measurements of advertising effectiveness. *Journal of Marketing 25*, pp. 59-62.

Leamer, E. E. (1992). Bayesian elicitation diagnostics. *Econometrica* 60(4), pp. 919-942.

Levitt, T. (1962). *Innovation in Marketing: New Perspectives for Profit and Growth*. McGraw-Hill.

Lichtenstein, S. (1978). Judged frequency of lethal events. *Journal of Experimental Psychology* 4, pp. 557-578.

Lindsley, O. R. (1962). A behavioral measure of television viewing. *Journal of Advertising Research, 2,* pp. 2–12.

Little, R. J. (2006). Calibrated Bayes: a Bayes/frequentist roadmap. *The American Statistician 60(3),* 213–223.

Loudon, D.L. and A.J. Della Bitta, (1983). *Consumer behavior: concepts and applications*. New York: McGraw-Hill.

Lunn, D. J., Thomas, A., Best, N. and Spiegelhalter, D. (2000). WinBUGS – A Bayesian modelling framework: concepts, structure and extensibility. *Statistics and Computing* 10, pp. 325–337.

Lutz, R. J. (1989). Positivism, Naturalism and Pluralism in Consumer Research: Paradigms and paradise. *Advances in Consumer Research 16,* pp.1-7.

Macdonald, R. R. (2002). The incompleteness of probability models and the resultant implications for theories of statistical inference. *Understanding Statistics, 1,* pp. 167–189.

Macklin, N. C. (1986). Classical conditioning affects in product/character pairings presented to children. In R. J. Lutz (Ed.), *Advances in consumer research* 13, pp. 198–203. Provo, UT: Association for Consumer Research.

Maddena, G. and Tanb, J. (2007). Forecasting telecommunications data with linear models. *Telecommunications Policy 31*, pp. 31–44

Mahajan, V., and Wind, J. (1992). New product models: Practice, shortcomings and desired improvements. *Journal of product innovation management, 9(2)*, pp. 128-139.

Malakoff, D. (1999). Bayes offers a 'new' way to make sense of numbers. *Science, 286(5444*), pp. 1460-1464.

Marsden, D. and Littler, D. (1998). Positioning alternative perspectives of consumer behaviour. *Journal of Marketing Management*, *14(1–3),* pp. 3–28.

Mason, J. (2002) *Qualitative Researching (Second Edition),* London: Sage.

May, T. 2001. *Social research – issues, methods and process (third edition)*. Open University press.

McCall, M. and Belmont, H. J. (1996). Credit card insignia and restaurant tipping: Evidence for an associative link. *Journal of Applied Psychology 81*, pp. 609–613.

McCracken, G. (1989). Who is the celebrity endorser? Cultural foundations of the endorsement process. *Journal of Consumer Research*, 16(3), pp. 310–321.

McGrayne S. B. (2012). *The theory that would not die: How Bayes' rule cracked the Enigma code, hunted down Russian submarines, and emerged triumphant from two centuries of controversy*. New Haven, CT: Yale University Press

McGuire, W.J. (1976a). Some internal psychological factors influencing consumer choice. *Journal of Consumer Research* 2, pp. 302-319.

McGuire, W.J. (1976b). 'An information processing model of advertising effectiveness'. In: H.L. Davis and A.J. Silk (eds.), Behavioral and management sciences in marketing. New York: Ronald Press.

McPhee, W. N. (1963). *Formal theories of mass behavior*. New York: The Free Press.

McQueen, J., Foley, C. and Deighton, J. (1993). Decomposing a brand's consumer franchise into buyer types. In: D. Aaker and A. Biel (eds.), Brand equity and advertising: Advertising's role in building strong brands, 235-246. Hillsdale, NJ: Lawrence Erlbaum Associates.

Milliman, R. E. (1982). Using background music to affect the behavior of supermarket shoppers. *Journal of Marketing* 46, pp. 86–91.

Moala, F. A. and O'Hagan, A. (2010). Elicitation of multivariate prior distributions: A nonparametric Bayesian approach. *Journal of Statistical Planning and Inference*, *140(7)*, pp. 1635-1655.

Montgomery, A. and Rossi, P. (1999). Estimating price elasticities with theory-based priors. *Journal of Marketing Research 36(4),* 413–423.

Morgan, R.P., 2000. A customer-oriented framework of brand equity and loyalty. *International Journal of Market Research 42(3),* pp. 65–120

Murphy, J. M. (1990). *Brand Strategy*. Director Books. Cambridge.

Nagle T. T. (1987). *The Strategy and Tactics of Pricing*. Englewood Cliffs, NJ: Prentice-Hall.

Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.

Newell, A. and Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Nogales, A. F., & Suarez, M. G. (2005). Shelf space management of private labels: a case study in Spanish retailing. *Journal of retailing and consumer services, 12(3),* pp. 205-216.

Nord, W.R. and Peter, J.P. (1980). A behavior modification perspective on marketing. *Journal of Marketing*, 44, pp. 36-47.

Oakley, J. E. and O'Hagan, A. (2005). Uncertainty in prior elicitations: a nonparametric approach. *Research Report No. 521/02 Department of Probability and Statistics*, University of Sheffield.

O'Hagan, A. (1994). *Kendall's advanced theory of statistics, volume 2b: Bayesian inference*. Cambridge: University press.

O'Hagan, A. (1998). Eliciting expert beliefs in substantial practical applications. *Journal of the Royal Statistical Society Series D* 47(1), pp. 21-35.

Oliveira-Castro, J. M., Foxall, G. R. and Schrezenmaier, T. C. (2005). Patterns of consumer response to retail price differentials. *Service Industries Journal* 25(3), pp. 309-335.

Oliveira-Castro, J. M., Foxall, G. R. and Schrezenmaier, T. C. (2006). Consumer brand choice: Individual and group analysis of demand elasticity. *Journal of Experimental Analysis of Behaviour* 85(2), pp. 147-166.

Oliveira-Castro , J. M. Foxall, G. R., James, V. K. B , Pohl, R. H. B. F., Dias, M. B. and Chang, S. W. (2008). Consumer-based brand equity and brand performance. *The Service Industries Journal* 28(4) pp. 445–461.

O'Shaughnessy, J. (1997). Temerarious directions for marketing. *European Journal of Marketing* 31(9/10), pp. 677–705.

Ouellette, J. A. and Wood , W. (1998). Habit and intention in everyday life: The multiple processes by which past behavior predicts future behavior. *Psychological Bulletin,* 124, pp. 54–74.

Perks W. (1947). Some observations on inverse probability including a new indifference rule. *Journal of the Institute of Actuaries, 73*, pp. 285-334.

Peter, J.P. and Nord, W.R. (1982). A clarification and extension of operant conditioning principles in marketing. *Journal of Marketing* 46, pp. 102-107.

Peter, J. P., and Olson, J. C. (1987). *Consumer behavior: Marketing strategy perspectives.* Homewood, IL: Richard D. Irwin.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing* 124, 125.

Poirier, D. J. (2006). The growth of Bayesian methods in statistics and economics since 1970. Bayesian Analysis 1(4), pp. 969-980.

Porto , R. B. and Oliveira-Castro, J. M. (2013). Say–Do Correspondence in Brand Choice : Interaction Effects of Past and Current Contingencies, *The Psychological Record* 63, pp. 345–362.

Punj, G. N., and Hillyer, C. L. (2004). A cognitive model of customer-based brand equity for frequently purchased products: Conceptual framework and empirical results. *Journal of Consumer Psychology, 14(1-2),* pp. 124-131.

Putnam, H. (2002). *The collapse of the fact/value dichotomy and other essays*. Harvard University Press, Cambridge, MA.

Rajala, A. K., and Hantula, D. A. (2000). Towards a behavioral ecology of consumption: Delay-reduction effects on foraging in a simulated Internet mall. *Managerial and Decision Economics* 21, pp. 145–158.

Razran, G. H. S. (1938). Conditioning away social bias by the luncheon technique. P*sychological Bulletin*, 37, 481.

Reber, A. S., Allen, R., and Reber, E. S. (2009). *The penguin dictionary of psychology 4th Ed)*. London: Penguin.

Reyes, P. M., and Frazier, G. V. (2007). Goal programming model for grocery shelf space allocation. *European Journal of Operational Research, 181(2),* pp. 634-644.

Robert, C. P., and Casella, G. (1999). The Metropolis—Hastings Algorithm. In *Monte Carlo Statistical Methods* (pp. 231-283). Springer New York.

Roberts A. (1980). The decision between above- and below-the-line. *Admap*, 588–92.

Rogers, A., Morgan, P. H. and Foxall, G. R. (2017). Building Consumer Understanding by Utilizing a Bayesian Hierarchical Structure within the Behavioral Perspective Model. *Behavioral Analyst,* In Print.

Romero, S., Foxall, G. R., Schrezenmaier, T. C., Oliveira-Castro, J. M. and James, V. (2006). Deviations from matching in consumer choice. *European Journal of Behavior analysis* 7(1), pp. 15-39.

Rossi, P. E. and Allenby, G. M. (1993). A Bayesian approach to estimating household parameters. *Journal of Marketing Research*, 171–182.

Rossi, P. E. and Allenby, G. M. (2003). Bayesian statistics and marketing. *Marketing Science 22(3),* 304–328.

Rothschild, M. L., and Gaidis, W. G. (1981). Behavioral learning theory: Its relevance to marketing and promotions. *Journal of Marketing*, 45, pp. 70–78.

Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *The Annals of Statistics 12*, 1151–1172.

Samaniego, F. J. and Reneau, D. M. (1994). Towards a reconciliation of the Bayesian and frequentist approaches to point estimation. *Journal of the American Statistical Association 89*, 947–957.

Saunders, M., Lewis, P. and Thornhill, A. (2009) *Research methods for business students (Fifth Edition)*, England: Pearson.

Senge, P. (1990). *The Fifth Discipline. The art and practice of the learning organisation*. Doubleday, New York.

Sharp, B., Wright, M. Dawes, J., Driesener, C., Meyer-Waarden, L., Stocchi, L. and Stern, P. (2012). It's a Dirichlet world. Modeling individuals' loyalties reveals how brands compete, grow, and decline. *Journal of Advertising Research*. pp. 203-213.

Sheth, J. N., Banwari, M. and Bruce, I. N. (1999). *Customer Behaviour Consumer Behaviour and Beyond.* The Dryden Press, Harcourt Brace College Publishers.

Shimp, T. A., Stuart, E. W. and Engle, R. W. (1991). A program of classical conditioning experiments testing variations in the conditioned stimulus and context. *Journal of Consumer Research*, 18(1), pp. 1–12.

Shocker, A.D. and Weitz, B. (1988). A perspective on brand equity principles and issues. In: L. Leuthesser, ed. Defining, measuring, and managing brand equity. Summary of March 1988 Conference, Austin, TX. Marketing Science Institute, Cambridge, MA, Report no.88-104.

Silverman, D. (2010) *Doing Qualitative Research (Third Edition)*, London: Sage.

Simon, C. J., and Sullivan, M. W. (1993). The measurement and determinants of brand equity: A financial approach. *Marketing science, 12(1)*, 28-52.

Skinner, B. F. (1938). *The behavior of organisms*. New York: Appleton-Century-Crofts.

Skinner, B. F. (1950). Are theories of learning necessary? *Psychological Review* 57, pp. 193-216.

Skinner, B. F. (1953). *Science and Human Behaviour.* New York: The Macmillan Company.

Skinner, B. F. (1957). V*erbal behaviour*. Englewood Cliffs, NJ: Prentice-Hall.

Skinner, B. F. (1969). *Contingencies of reinforcement: a theoretical analysis*. New York: Appleton-Century-Crofts.

Skinner, B.F. (1972). *Beyond freedom and dignity*. London: Cape.

Skinner, B.F. (1974). *About behaviourism*. London: Cape.

Skinner, B. F. (1985). Cognitive science and behaviourism. *British Journal of Psychology* 76, pp. 291-301.

Smith, C. L. and Hantula, D. A. (2003). Pricing effects on foraging in a simulated Internet shopping mall. *Journal of Economic Psychology* 24(5), pp. 653–674.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society B* 64(4), *pp.* 583–639.

Stent, G. S. (1975). Limits to the scientific understanding of man. *Science*, 187, pp. 1052-1057.

Stephens, D. W. and Krebs, J. R. (1986). *Foraging theory*. Princeton, NJ: Princeton University Press. Stratton, J

Stobart, P. (1990). “Brands—The Way to the Yellow Brick Road?”. *Accountancy, May*, pp. 27–28.

Studer, R.G. (1973). “Man-environment relations: discovery or design", in Preiser, W.F.E. (Ed.), Environmental Design Research, Vol. 2, Dowden, Hutchinson and Ross, Stroudsburg, PA, pp. 136-51.

Style, C., and Ambler, T. (2000). The impact of relational variable on export performance: an empirical investigation in Australia and the UK. *Australian Journal of Management, 25(2)*, pp. 261-281.

Taylor, S. A., Celuch, K. and Goodwin, S. (2004). The importance of brand equity to customer loyalty. *Journal of Product and Brand Management* 13(4), pp. 217-227.

Telser, L. G. (1962). The demand for branded goods as estimated from consumer panel data. *The Review of Economics and Statistics*, 300–324.

Thomas, A.B. (2004) *Research Skills for Management Studies*, Abingdon: Routledge.

Tierney, L. (1994). Markov Chains for Exploring Posterior Distributions. *The Annals of Statistics 22(4)*, 1701–1728.

Tiffany, S. T. (1990). A cognitive model of drug urges and drug-use behavior: Role of automatic and nonautomatic processes. *Psychological Review*, 97(2), Apr, pp. 147-168.

Till, B. D., Stanley, S. M. and Priluck, R. (2008). Classical conditioning and celebrity endorsers: An examination of belongingness and resistance to extinction. *Psychology & Marketing* 25(2), pp. 179–196.

Turley, L. W., and Milliman, R. E. (2000). Atmospheric effects on shopping behavior: a review of the experimental evidence. *Journal of business research, 49(2)*, pp. 193-211.

Viscusi , W. K. (1985). Are individuals Bayesian decision makers? *The American Economic Review , Papers and Proceedings of the Ninety-Seventh Annual Meeting of the American Economic Association* 75(2), pp. 381-385.

Wallace, T. D. Weaker criteria and tests for linear restrictions in regression, *Econometrica, 40*, pp. 689-98.

Watkins, T. (1986). *The economics of the brand: A marketing analysis.* McGraw-Hill Book Company.

Watson, J. B. (1930). *Behaviorism (Revised edition)*. Chicago: University of Chicago Press

Webster, F. E. Jr. (1994). *Market-Driven Management.* New York: Wiley.

Wells, V. K. (2014). Behavioural psychology, marketing and consumer behaviour: a literature review and future research agenda. *Journal of Marketing Management*, 30(11–12), pp. 1119–1158.

Welp, M., de la Vega-Leinert, A., Stoll-Kleemann, S. and Jaeger, C. C. (2006). Science-based stakeholder dialogues: Theories and tools. *Global Environmental Change, 16(2)*, pp. 170-181.

Wentz, L. and Martin, G. (1989). How experts value brands. *Advertising Age, 16*, p. 24

Wicker, A. W. (1969). Attitude versus actions: the relationship of verbal and overt behavioural responses to attitude objects, *Social Issues*, 25, pp. 41-78.

Wicker, A.W. (1987). ``Behavior settings reconsidered: temporal stages, resources, internal dynamics, context", in Stokols, D. and Altman, I. (Eds), Handbook of Environmental Psychology, John Wiley and Sons, New York, NY, pp. 613-53.

Wohlwill, J.F. (1973). ``The environment is not in the head", in Preiser, W.F.E. (Ed.), Environmental Design Research, Vol. 2, Dowden, Hutchinson and Ross, Stroudsburg, PA, pp. 166-81.

Wood, L. M. (1995). Brands: the asset test. *Journal of Marketing Management*, *11(6)*, pp. 547-570.

Wood, L. (2000). Brands and brand equity: definition and management, *Management Decision* 38(9), pp. 662-668.

Wright, M. and Sharp, B. (2001). The effect of a new brand entrant on a market. *Journal of Empirical Generalisations in Marketing Science* 6, pp. 15-29.

Xu, X., Qi, Y. and Hua, Z. (2010). Forecasting demand of commodities after natural disasters. *Expert Systems with Applications 37*, pp. 4313–4317.

Yan, J. (2012). *Essential value and consumers' brand choice* (Unpublished doctoral dissertation). Cardiff University.

Yang, H. Y., Hao Ye, H., Wang, G., Khan, J. and Hu, T. (2006). Fuzzy neural very-short-term load forecasting based on chaotic dynamics reconstruction. *Chaos, Solitons and Fractals 29*, pp. 462–469

Appendix

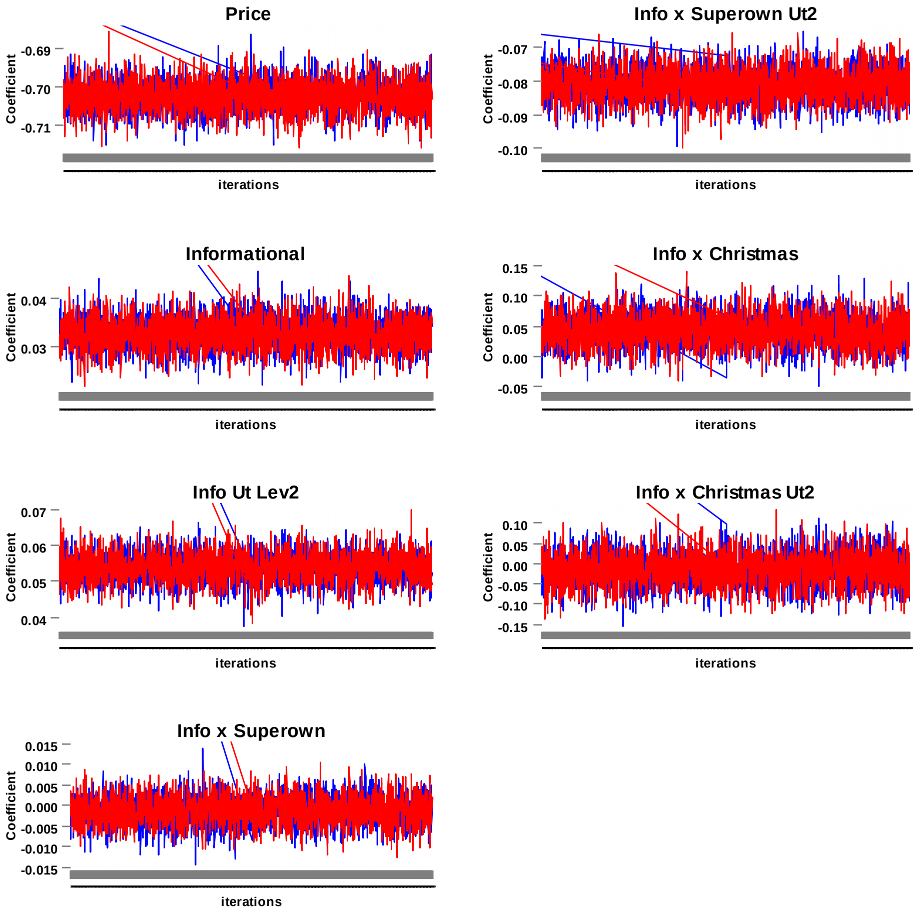

Figure 111: Biscuits Hierarchical

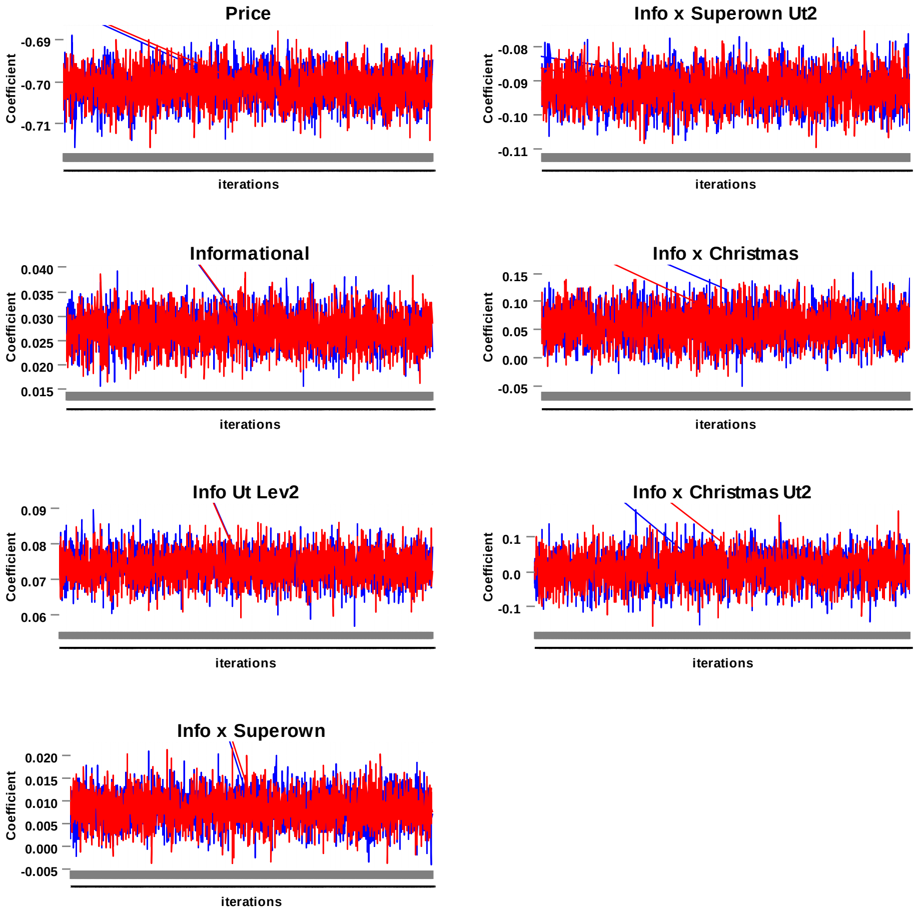

Figure 112: Biscuits Non Hierarchical

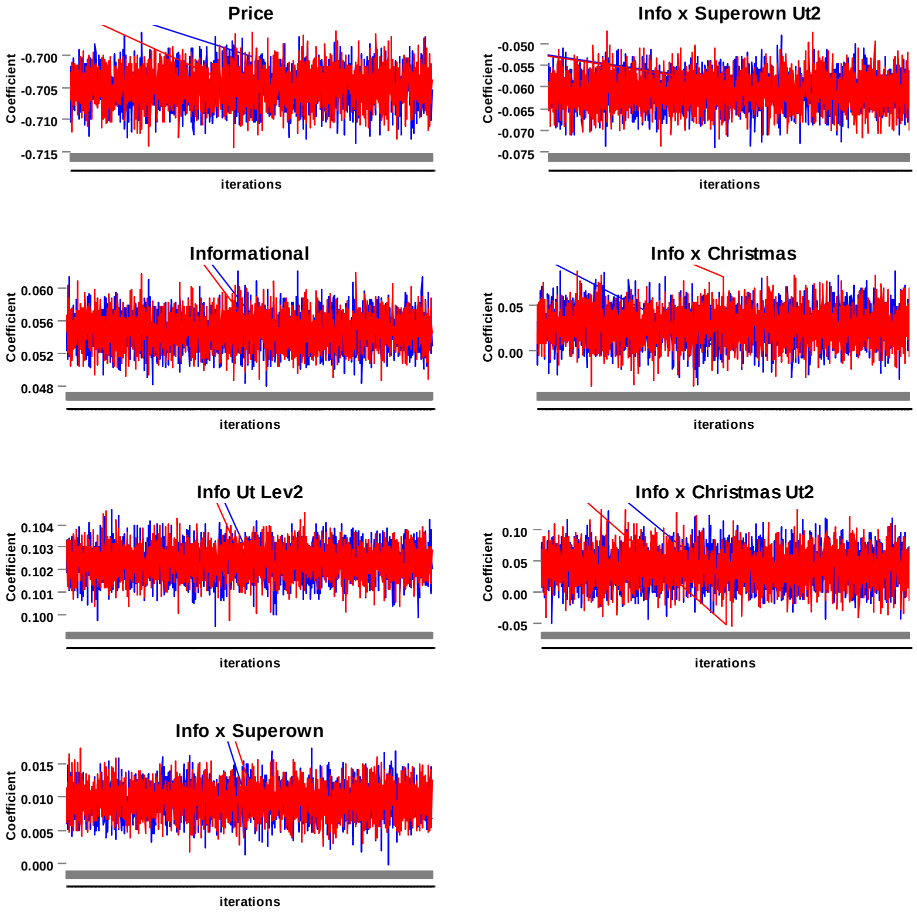

Figure 113: Biscuits Hierarchical Informative

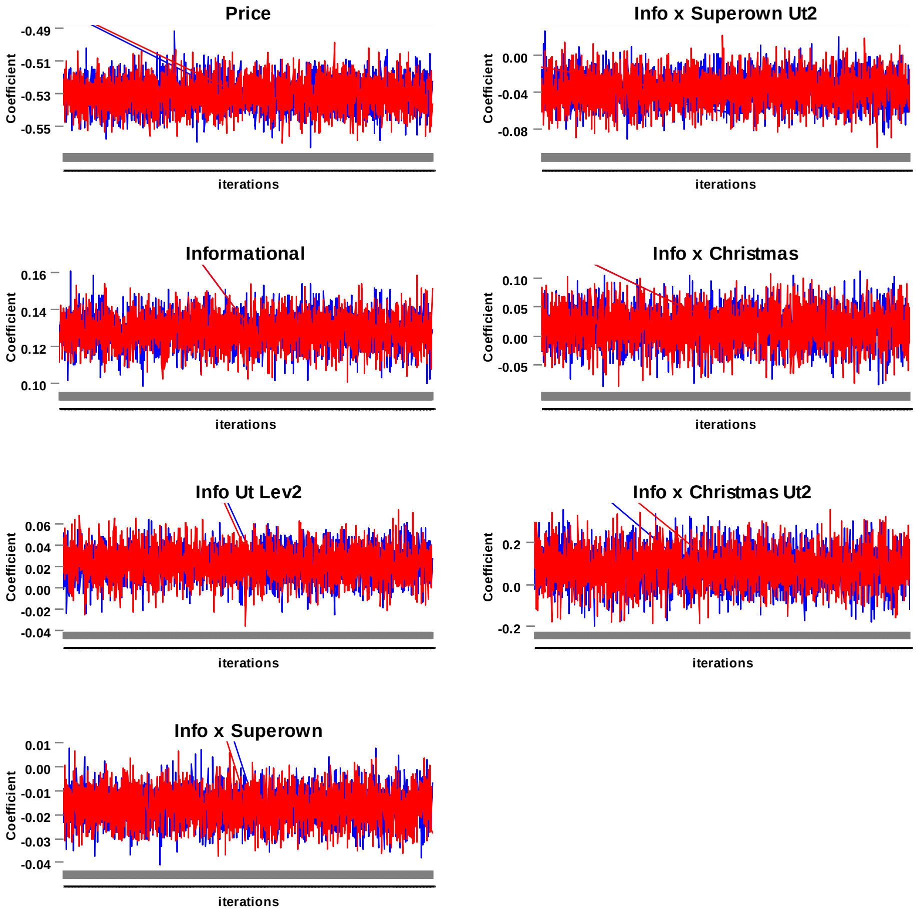

Figure 114: Fruit Juice Hierarchical

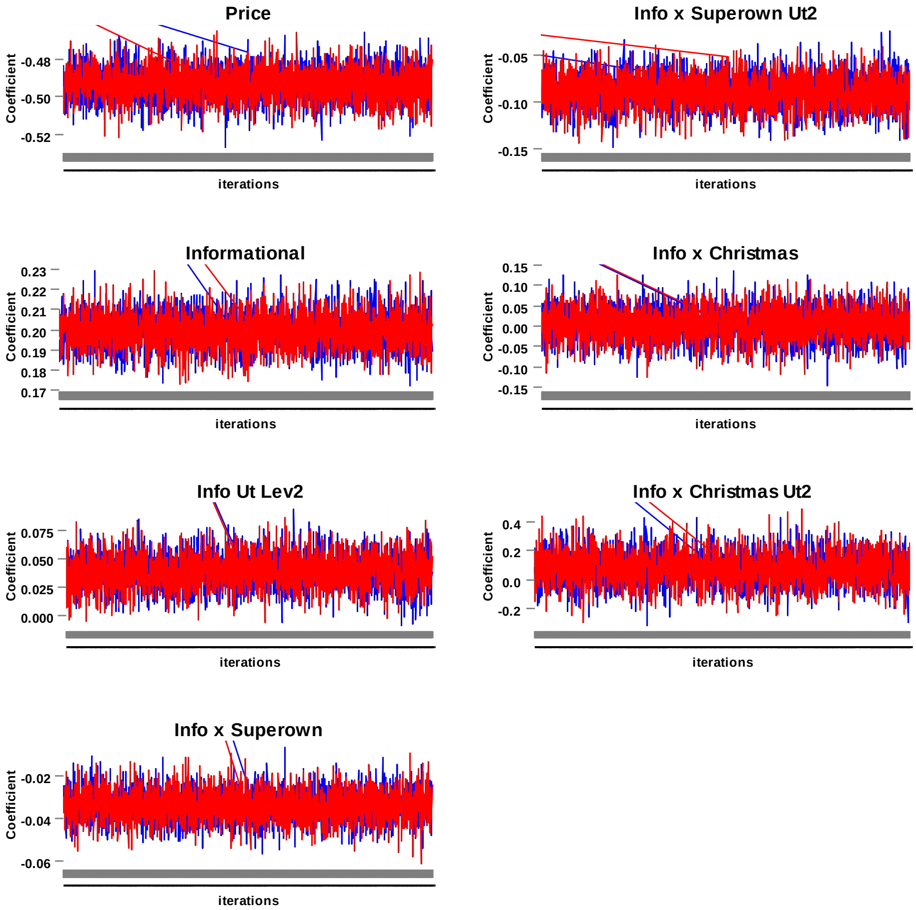

Figure 115: Fruit Juice Non Hierarchical

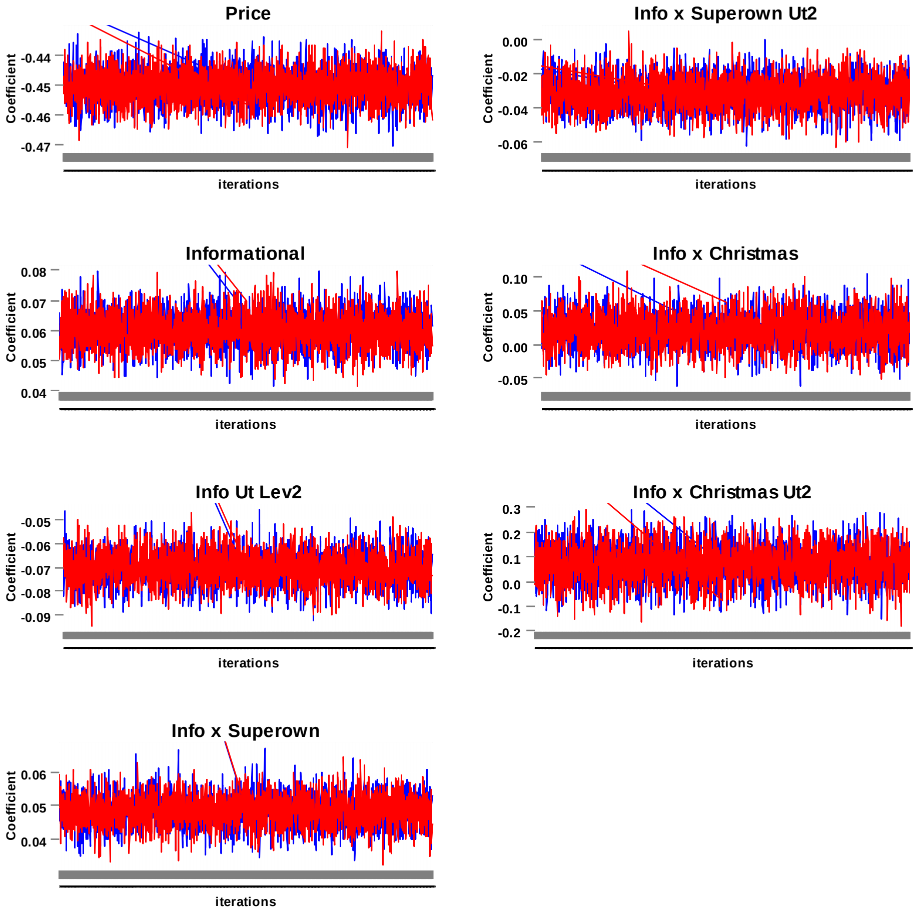

Figure 116: Fruit Juice Hierarchical Informative

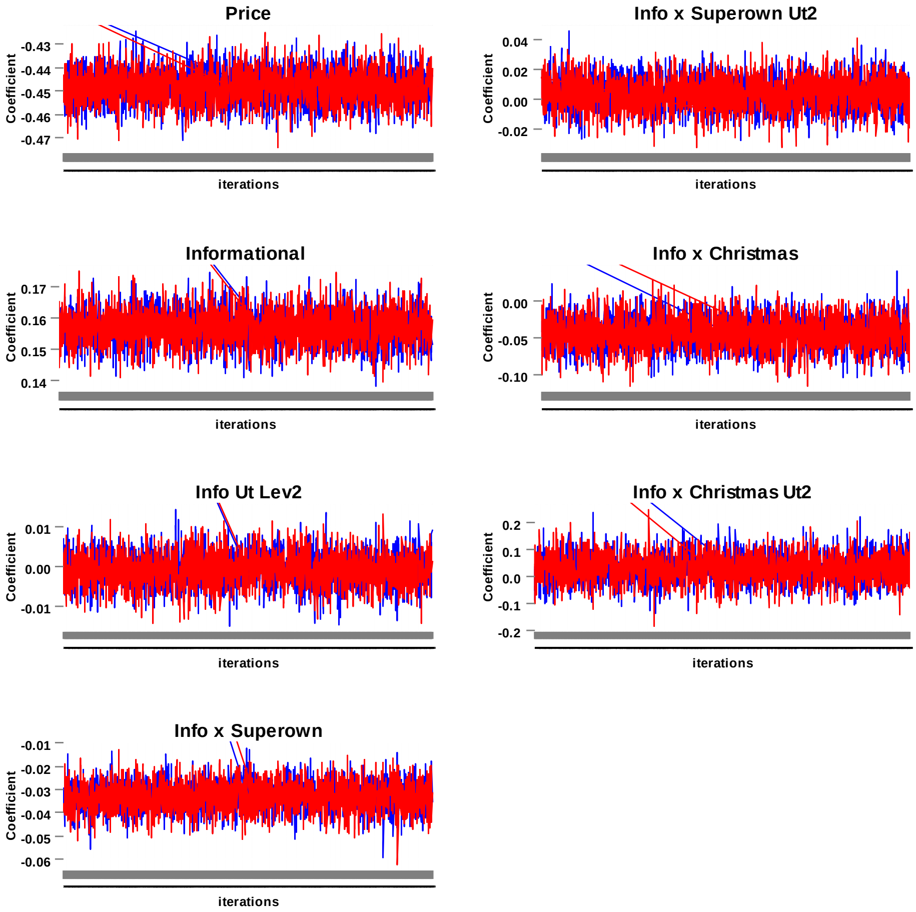

Figure 117: Yellow Fats Hierarchical

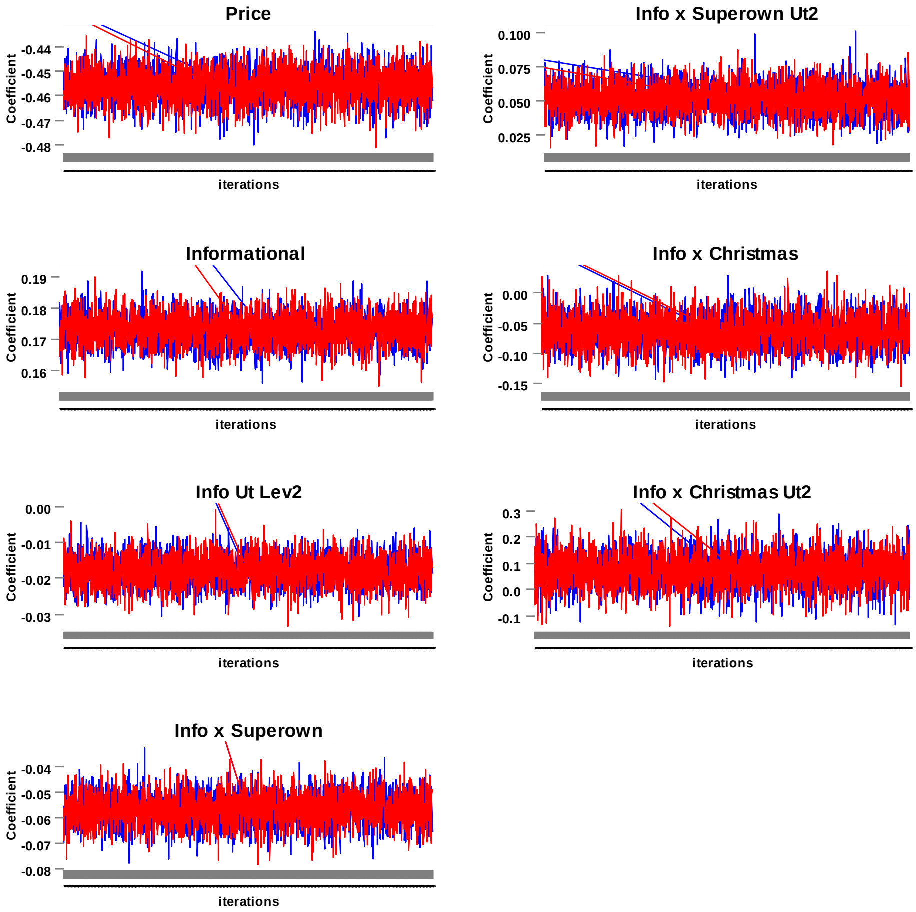

Figure 118: Yellow Fats Non Hierarchical

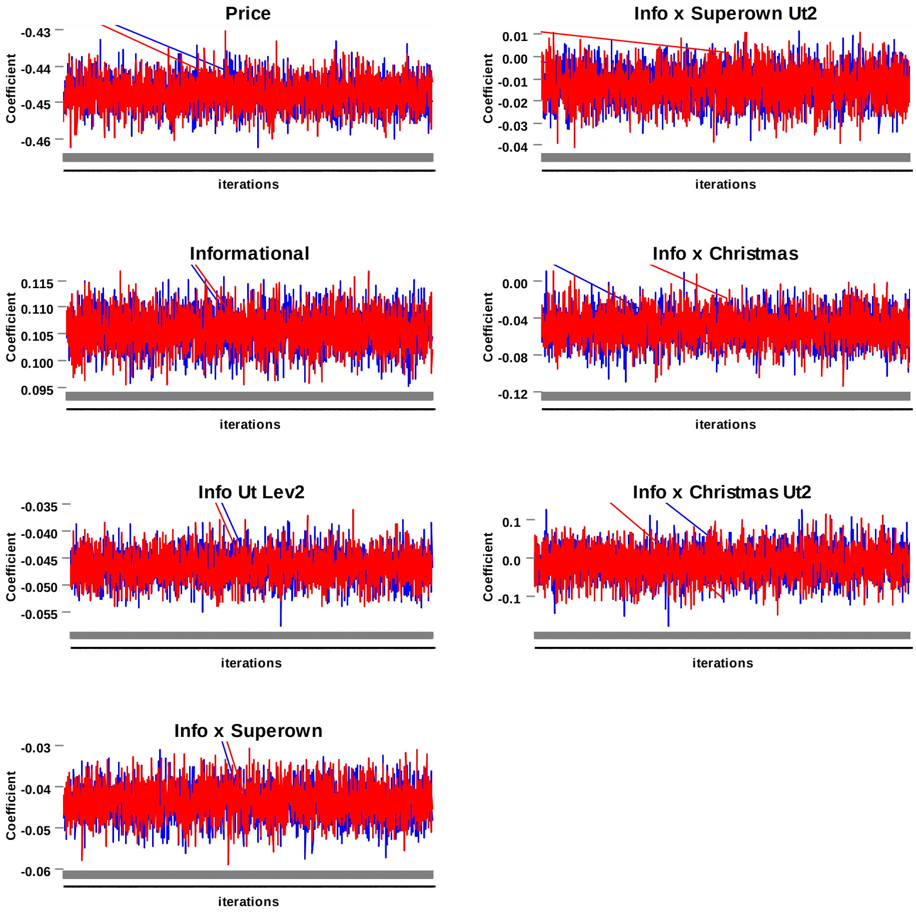

Figure 119: Yellow Fats Hierarchical Informative

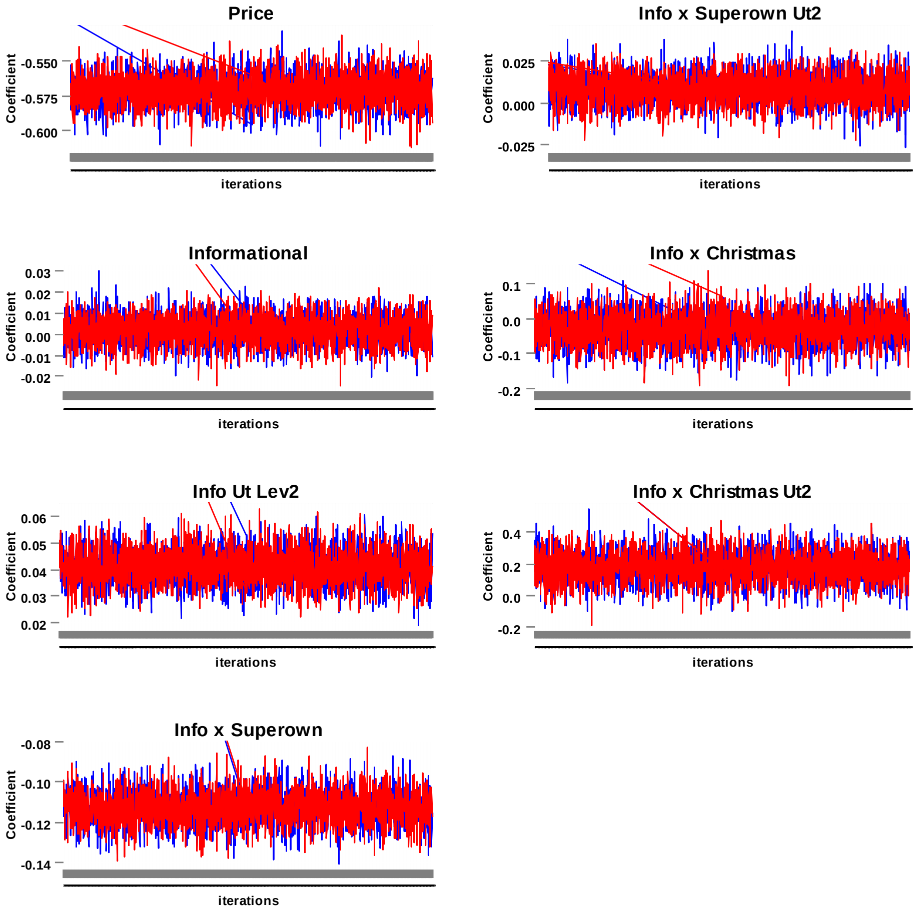

Figure 120: Beans Hierarchical

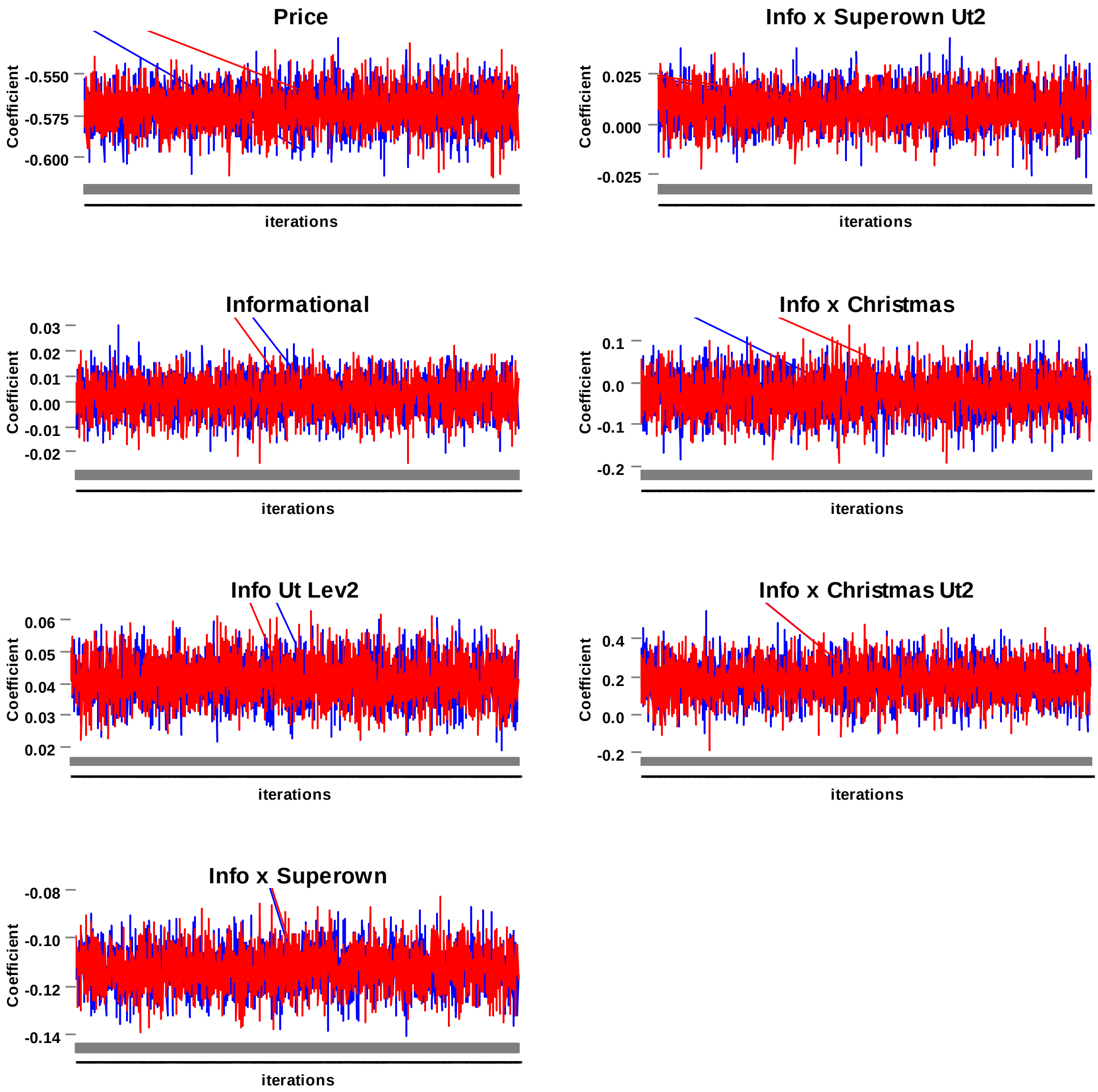

Figure 121: Beans Non Hierarchical

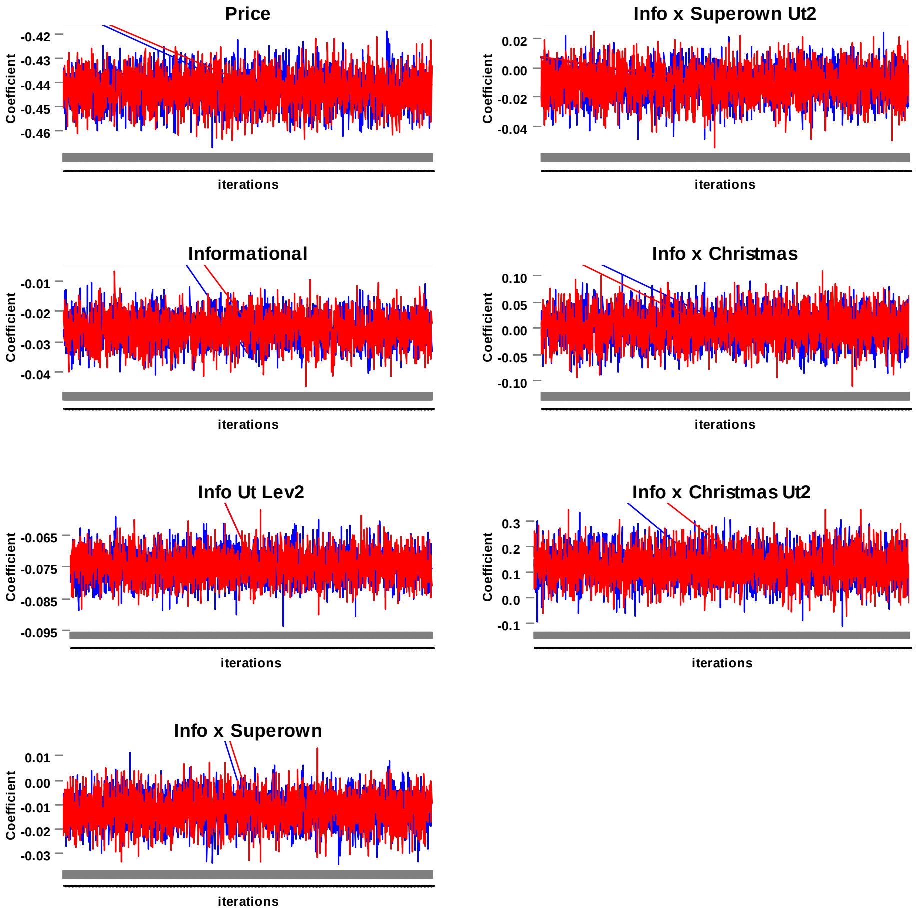

Figure 122: Beans Hierarchical Informative